Centre for Economic Performance

Discussion Paper

ISSN 2042-2695

No. 1904
March 2023

Who stands on the shoulders of Chinese (scientific) giants? Evidence from chemistry

Shumin Qiu
Claudia Steinwender
Pierre Azoulay

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Economic and Social Research Council

Abstract

In recent decades, Chinese researchers have become preeminent contributors to the scientific enterprise, as reflected by the number of publications originating from Chinese research institutions. China's rise in science has the potential to push forward the global frontier, but mere production of knowledge does not guarantee that others are able to build on it. In this manuscript, we study how fertile Chinese research is, as measured by citations. Using publication and citation data for elite Chemistry researchers, we show that Chinese authored articles receive only half the citations from the US compared to articles from other countries. We show that even after carefully controlling for the "quality" of Chinese research, Chinese PIs' articles receive 28% fewer citations from US researchers. Our results imply that US researchers do not build as readily on the work of Chinese researchers, relative to the work of other foreign scientists, even in a setting where Chinese scientists have long excelled.

Key words: research and development, international spillovers, economics of science, citations, patent citations
JEL: I23; O30;O35

This paper was produced as part of the Centre's Growth Programmes. The Centre for Economic Performance is financed by the Economic and Social Research Council.

We thank Richard Freeman, Alessandra Allocca, Pat Kline, as well as seminar audiences at USC, UC Berkeley, the CRC, and the NBER China Working Group for insightful comments. We thank Kimberly White, Tabea Rudolf, and Lynn van den Busch for outstanding research assistance.

Shumin Qiu, East China University of Science and Technology. Claudia Steinwender, Ludwig Maximilian University, NBER and Centre for Economic Performance, London School of Economics. Pierre Azoulay, MIT Sloan School of Management and NBER.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

1 Introduction

China has overtaken the United States to become the world's largest producer of scientific publications (Tollefson, 2018; Xie and Freeman, 2019). Even with the acknowledgement that (scientific) quantity might have a quality of its own, interpreting this dramatic increase is difficult. From the standpoint of its impact on the global economy, an important question is whether, beyond its undeniable quantitative importance, Chinese research contributes to pushing the world scientific frontier outward.

Recent empirical findings lend credence to the view that the quality of Chinese research has improved in concert with the number of articles emanating from Chinese research institutions (and researchers). For instance, the incidence of Chinese addresses (and Chinese names) in world-leading journals such as *Science* and *Nature* has more than doubled between 2000 and 2016 (Xie and Freeman, 2020). The average number of citations per article, and China's overall share of citations has also risen markedly (Xie and Freeman, 2019).

These stylized facts notwithstanding, the extent to which Chinese scientific knowledge offers "broad shoulders" for follow-on researchers to stand on remains an open question. In particular, how are citations to Chinese research geographically distributed? The last twenty years have seen a 2.5-fold increase in the number of Chinese academic scientists (PRC National Bureau of Statistics, various years), many of them working in relatively new, less research-intensive institutions. Because of this increase in scientific labor supply, the rising impact of Chinese research could merely reflect an elevated propensity on the part of Chinese researchers to cite research "made in China" (Qiu et al., 2021). This could arise because of the localized nature of knowledge spillovers, or because of other frictions, such as lower communication costs for researchers who share the same language (Xie and Freeman, 2020). Conversely, foreign scientists might discount the importance of Chinese scholarship, compared to research produced in their home countries or elsewhere in the world.

Contrasting Chinese and non-Chinese (and non-US) researchers, we study the extent to which articles of similar observable quality are differentially likely to be cited by researchers based in the US. Our preferred specifications point to a "China citation discount" equal to 28% of the baseline probability of citation. This discount is halved for Chinese researchers who received at least some of their scientific training in the United States, and not present for US citers with Chinese names. We also find evidence that this discount is not a mere reflection of clustering of Chinese researchers in particular subfields that are less likely to be cited by US scientists. Nor is it likely to reflect ethnic animus, since we do not observe a similar discount for researchers with Chinese names located outside China. In addition, a similar discount appears present in US citations to the scientific literature contained in patents.

The China discount has been stable over the past two decades. Among the top Chemistry nations (by number of publications in our sample), no other country experiences a citation discount from the US; instead, Switzerland and Germany experience small citation premia. These results are notable because our choice of setting—elite scientists, in a domain where China has a long tradition of excellence—would seem to be one without particular impediments to the diffusion of knowledge across borders. Yet this appears to be far from the case. Together, our results imply that China's pronounced citation "home bias" reflects, at least in part, missing citations from non-Chinese authors, perhaps offset by a surplus of citations attributable to the vastly expanded pool of Chinese potential citers when aggregating citation data at the level of a field, a journal, or an entire country.

The manuscript proceeds as follows. We begin by a brief history of chemistry research in China in section 2. Section 3 describes our data sources and sample. Section 4 provides evidence of a Chinese citation discount using aggregate data, while section 5 highlights the matched article pair design that will help us establish that the discount survives our attempts to hold research "quality" constant. Section 6 and section 7 reports the results of the analysis respectively for citations contained in articles as well as patents. Section 8 concludes.

2 Chinese Research in Chemistry

For thousands of years ancient China led the world with remarkable inventions and achievements in the chemical and metallurgical arts (Agnew, 1997). Many important empirical discoveries originated from ancient Chinese alchemy and medicinal Chemistry; their translation into Western languages had a pronounced influence on modern chemical science (Leicester, 1971). For example, many historians believe that gunpowder technology, one of the most influential inventions in human history, had its origins in China (581-681 CE), and then spread to the Middle East and Europe along the Silk Road (Needham et al., 1986). Chinese pre-modern "scientists" also pioneered the manufacturing processes for salt, wine, paper, and porcelain (Li, 1948; **?**).

Although historians suggest that modern Chemistry grew, at least in part, out of the work of Chinese alchemists (Leicester, 1971), Chemistry as a modern science was absent in China until the 19^{th} century, when European science was introduced through missions, trade, and wars (Li, 1948). In the late Ch'ing Dynasty (mid-to-late 19^{th} century), during which the rulers adopted a closed-door policy with very limited communication with the outside world, Chinese Chemistry (as well as other sciences) lagged far behind western countries. After wars with European countries broke open China's door, modern Chemistry started to develop with the purpose of "learning from foreigners to compete with them" as China became integrated into the global "Republic of Science" (Bai, 2000). Research by western chemists were intensively translated into Chinese and disseminated in China: Between 1912 and 1949, 41% of Chemistry articles and textbooks were translated from English, while the remainder (very few of them were original scientific research) were written by Chinese chemists [1].

In order to further acquire frontier knowledge, a first wave of Chinese students were sent to the United States for scientific training under the aegis of the Boxer Indemnity Scholarship Program. The first generation of returnee students had a lasting influence on Chinese modern science and some of them became pioneers and academic leaders in the field of Chemistry after coming back to China.[2]

The rapid development of modern Chinese Chemistry took place after the founding of the People's Republic of China, especially after the deep opening policy begun in 1978 (Bai, 2000). Between 2000 and 2017, the number of Chinese universities increased by 140% (from 753 to 1,805), and correspondingly the number of Chemistry departments rose by 182% (from 243 to 686). Research faculty in Chinese universities increased by 69% during the same period, and the number of Chemistry researchers tripled. Public research funding invested in Chemistry also shows a 14-fold increase between 2000 and 2017, higher than the ten-fold increase observed for other fields on average.[3]

Meanwhile, China has continuously expanded global collaboration and communication by funding students' graduate studies abroad, and facilitating Chinese scholars' participation in international collaboration through the funding of shorter-term stays in frontier countries. The number of state-financed students studying abroad increased five-fold, from 7,564 in

---

[1] Source: National bibliography of the Republican period (1911-1949) produced by Beijing Library.

[2] In 1908, US and China reached an agreement to use the excess funds from the Boxer Indemnity to establish a scholarship program for Chinese students to study in the US. During 1909 to 1929, this program sent around 1,300 Chinese students to the US, studying in selected fields that served the urgent needs of Chinese development, including science, engineering, medicine, and agriculture. Celebrated alumni of this program include Hou Debang [BA, MIT, MS, Columbia], Chen Hwang [BS, MS, MIT], Chang Tsun [BA, MIT], Hsu Paul Hwang [BS, MIT], and Chien Shih-Liang [MS, PhD, UIUC]).

[3] The source for these figures, and those mentioned below is the *Compilation of University Science and Technology Statistics* produced by the Chinese Ministry of Education.

2000 to 46,347 in 2017, whereas attendance of international conferences increased almost eight-fold during the same period. Between 1978 and 2018, a total of 5.86 million students studied abroad, 82% of whom returned to China. The flow of transnational human capital, particularly the return of elite scientists, has helped create a solid foundation for Chinese scientific research. Scientists holding overseas degrees account for 37% of the total number of members of Chinese Academies of Sciences and Engineering elected between 1955 to 2009. During this period, 300 US-trained academics returned to China, a figure to be compared with 160 Soviet-trained and 80 UK-trained academics who returned during the same period.[4] In Chemistry specifically, there is evidence that students receiving graduate training in the United States are among the best and brightest. Gaulé and Piacentini (2013) document that Chinese students perform about as well as the awardees of the prestigious NSF doctoral fellowship program, and far better than other foreign students.

While China has been a rising country across a broad cross-section of scientific domains, its status as a producer of frontier scientific knowledge has stood out in a narrower set of fields, Chemistry preeminent among them. China's share of world publications in *Web of Science* has grown from 5.33% in 2000 to 25.94% in 2018, with Chemistry, Engineering and Materials Science being the strongest contributors to growth, as can be seen in Figure 1. According to *Nature Index*, a database consisting of research articles published in an independently selected group of high-quality science journals, China's fractional count of articles grew by 84% between 2012 and 2017, making the country second only to the United States. In some Chemistry subfields, such as organic chemistry, China even surpassed the United States in recent years to become the world's top producer of publications.

Table 1 demonstrates the importance of Chinese elite researchers. According to the annual *Highly Cited Researchers* (HCRs) rankings published by *Clarivate Analytics* between 2014 and 2018, Chemistry ranks highest among scientific fields in terms of highly cited researchers (column 1). These 211 researchers account for 19.27% of the world's HCRs in Chemistry (column 2). Table 2 shows that Chinese chemists have become world-leading contributors compared to other countries. During 2000 to 2015, China's share of publications in Chemistry was 14.96%, ranking it second only after the United States. Japan is a distant third with 7.66%, followed by Germany, India, and the United Kingdom. The ranking with respect to HCRs is similar, with the United States accounting for the largest share of the world's elite chemists (43.01%), followed by China (19.27%).

---

[4] Source: *Survey Report of Academicians of Chinese Academy of Sciences and Chinese Academy of Sciences.*

China's strong position in Chemistry becomes particularly striking when we compare it to China's ranking with respect to all fields. Across the sciences, the United States are clearly the most dominant nation with 8,306 HCRs amounting to almost half (46.48%) of the world's top scientists. Second by a large margin is the UK with 1,701 HCRs (9.52% of the world), and China is third with 1,104 HCRs (6.18% of the world). This makes China's HCR share in Chemistry more than three times larger than the average across fields.

3 Description of Data Sources

The goal of this manuscript is to investigate how research undertaken in China disseminates compared to research undertaken by other countries.We focus on the publications of the world's best researchers, understood to be those with the highest rate of publications in a defined list of Chemistry journals. We focus on elite chemists from all countries, excluding the United States.[5]

Generating a list of elite chemists. We compile a list of 31 most impactful journals in the field of Chemistry.[6] This yields a sample of 31 journals that we label as "elite journals". We consider all original research articles in these 31 elite journals published between 2000-2018, the period of China's rise in science. We drop articles produced by teams larger than 15 co-authors. This yields a total number of 552,933 articles in elite journals.

Identifying elite researchers. Based on the author disambiguation work of Torvik and Smalheiser (2009) and its update to the 2018 version of *PubMed*, we are able to assign each article to unique authors. We focus on last authorship position, which indicates principal investigatorship according to the publication norms in the field of Chemistry. From a set of 124,966 unique last authors in the 552,933 articles, we select the top 1% in terms of the number of elite journal publications, and obtain a sample of 1,250 investigators—our group of elite scientists.

Researcher-level data. We focus on investigators from all countries, excluding the United States, which leaves us with 775 investigators. In order to obtain individual characteristics for them, we contacted each scientist through e-mail requests. In this process, we received 254 responses (32.77% response rate). For those who do not respond, we collected information

---

[5] As explained in more detail below, we will consider the US a large frontier country whose researchers are at risk of citing articles written by Chinese and non-Chinese scientists.

[6] The list displayed in Appendix Table B.1 reflects a number of *ad hoc* adjustments (based on consultations with a number of academic chemists) made to the list that we would have obtained had we used 2020 journal-impact factors to generate it.

through their laboratory or institution faculty page, their *Who's Who* profile, and Google searches. This process yielded sufficient information for an additional 497 researchers. 24 investigators were dropped because we couldn't obtain sufficient data for them from any of the sources. Our final sample thus consists of 751 investigators.

From these responses, we extract information about demographics (birth year, gender), PhD education (university, country, completion year), post-doctoral experience (organization and time period), as well as employment spells since post-doc (organization, country, and time period). We define the "year of independence" of each researcher as the year of their first faculty employment after post-doctoral education. We use the country that is associated with their most frequent affiliation on publications after career independence to assign each scientists to a unique country—this does not necessarily correspond to the nationality of the scientist.[7] Among the 751 top 1% most published star scientist 156 (20.78%) are from China, and 595 (79.22%) are from the rest of the world. Overall, the majority of scientists are male (96%), and 80.82% of scientists have some postdoctoral experience. Their average doctoral degree year is 1988, while the average year of independence is 1992, and the average number of post-doctoral years is 4. 11.32% of scientists hold a PhD degree from universities located in the United States, and 49.13% of them spent their post-doc years at institutions in the United States.

Publication data. We compile the full publication list for all 751 scientists in our sample published between the years 2000 and 2018.[8] To ensure that we capture only research that was influenced to a significant extent by the scientist, we restrict the publications in two ways. First, we focus on publications which list the scientist as last author, which indicates principal investigatorship according to publication norms prevalent in Chemistry. Second, we consider only articles that were published after the PI became an "independent" researcher, according to the definition above. Overall, our sample comprises 78,541 scientific articles in

---

[7] For robustness, we also assigned scientists to the country in which their first non post-doctoral position is located, with little change to the results. Scientists may move to different countries over the course of their career, and we worry about the selectivity bias that may arise from assigning each publication to the country in which it was produced, since opportunities to move could plausibly be related to productivity. In practice, this distinction does not matter very much: For 95% of publications, the country of the first faculty employment is the same as the country of the affiliation at the time of publication, and our results are robust to this alternative definition.

[8] We restrict our sample to publications after 2000 (because there were very few Chinese scientists actively publishing in international journals before), and before to 2018, to give every article in the sample at least three years to be cited (until 2021).

Chemistry.[9] On average, each scientist published 104.58 articles as last author in the time period we consider (*s.d.* = 58.33), ranging from a minimum of 45 to a maximum of 668 publications.

Citation data. We compile a list of citations of the publications to our elite researcher sample from *Web of Science*. Since we want to link citations to countries, we remove citing articles lacking country information (4.2% of citations), which results in our database comprising of 2,839,144 citation records from 2000 to 2021 for the 78,541 last-authored articles. Each article in our dataset received on average 36.18 cites. To uncover the causes of differences in cross-country citation behavior, we focus on the propensity of US researchers to cite articles that originate from China versus other countries. We single out the US, because it is undoubtedly a frontier country in Chemistry research that attracts collaborations and trainees from the world at large. Furthermore, its large size implies that citation linkages between the US and other countries are frequent enough to make the statistical analysis tractable. In order to ensure that we can unambiguously interpret cross-border citation linkages, we restrict the sample of citations to articles in which all authors are affiliated with a US institution. This yields 271,194 citations records for the 78,541 focal articles, belonging to 98,915 unique citing articles from the US.

4 US citations of Chinese articles: Aggregate evidence

We begin with the full set of articles that was published in the field of Chemistry between 2000 and 2018 (i.e., without restricting to elite researchers) and ask whether there is any difference in the number of citations Chinese articles receive from the US compared to articles written in other, non-US countries. In order to simplify the analysis, we restrict the set of citations to articles that can uniquely be assigned to specific countries, i.e., articles that list institutions from only one country, and estimate the following Poisson model via quasi-maximum likelihood:

$$E\left[US\_citations_i | X_i\right] = \exp\left(\beta_0 + \beta_1 China_i + \beta_2^{'} X_i\right)$$

---

[9] 71.11% of these articles are published in one of the 31 elite Chemistry journals listed in Appendix Table B.1, whereas 28.89% are published in other journals, including prestigious multidiciplinary outlets such as *Science*, *Nature*, or the *Proceedings of the National Academy of Sciences*.

where $US\_citations_i$ counts the number of citations received by article i, $China_i$ is an indicator variable for whether article i's last author is affiliated with a Chinese institution, and X_i is a vector of control covariates.

Chinese PIs receive on average 0.94 cites from US articles. This is only half of the cites that non-Chinese PIs receive from the US: 1.92. The Poisson regression estimates in column 1 of Table 3 reflects this fact and shows that a Chinese article receives on average 48% fewer US citations compared to an article from other non-US countries.[10] Because citations increase over time, and the rise of Chinese research has been more recent, we control in column 2 for publication year fixed effects, which reduces the effect to -34%. This confirms that Chinese research is not as readily cited by US articles than research from other countries.

One obvious explanation for this apparent citation discount may be that Chinese research is of lower quality than that undertaken in other countries. We implement a first attempt to control for the quality of articles by adding journal fixed effects in column 3 of Table 3, which reduces the citation discount further to 24%. Of course, controlling for journal quality may not be enough, as there remains large variation *within journals* with respect to how much follow-on research articles can inspire. It may well be the case that once we properly control for the quality of the research, the discount vanishes. This is what we examine in the remainder of the manuscript.

5 Empirical Strategy

To detect whether the Chinese citation discount can be explained by researchers' and articles' observable characteristics, we pursue the following strategy: First, we use coarsened exact matching (CEM) to find articles of comparable quality for each publication by elite Chinese PIs. Second, we define the risk set of articles that may be citing the elite researchers' publications based on topical relatedness. Finally, we estimate a linear probability model to identify the relative "fertility" of Chinese-authored articles, relative to those authored by non-Chinese PIs.

5.1 Matching Chinese with non-Chinese articles

In order to control more fully for the "quality" of each article, we match each publication by a Chinese scientist to a similar publication by a non-Chinese, non-US scientist. We refer to publications of Chinese PIs in our sample as the treatment group. For each of these

---

[10] Since $\exp(-0.662) - 1 = -0.48$. See the last row of Table 3 for equivalent % changes.

articles, we try to find at least one similar article authored by a non-Chinese, non-US PI, the control group. We implement a "Coarsened Exact Matching" procedure (Blackwell et al., 2009). The first step is to select a relatively small set of covariates on which we need to guarantee balance *ex ante*. This choice entails judgement, but is strongly guided by our desire to hold the "fertility" of cited articles approximately equal across the treatment and control groups. The second step is to create a large number of strata to cover the entire support of the joint distribution of the covariates selected in the previous step. In a third step, each observation is allocated to a unique strata, and for each observation in the treated group, control observations are selected from the same strata.

The procedure is coarse because we do not attempt to precisely match on covariate values; rather, we divide the support of the joint distribution of the covariates into a finite number of strata, and we match a treated observation if and only if a control observation can be recruited from this strata. An important advantage of CEM is that the analyst can guarantee the degree of covariate balance *ex ante*, but this comes at a cost: the more fine-grained the partition of the support for the joint distribution (i.e., the higher the number of strata), the larger the number of unmatched treated observations.

Proxy measures for article quality. We want to ensure that we compare articles that have the same 'quality.' The literature often uses citations to measure quality. However, we face two restrictions: First, we cannot use citations from US scientists as these correspond to our outcome variable. Second, in a companion paper we show that citations often exhibit strong 'home bias': articles are disproportionately cited by scientists from the same country (Qiu et al., 2021). If this home bias is especially pronounced in China, maybe because of an elevated propensity on the part of Chinese researchers to cite research 'made in China', this could imply that citations by Chinese authors are inflated, making them less of a valid proxy for article quality, relative to articles published by researchers located in other countries.

Indeed, the citations that China receives from itself as a share of all citations it receives from the world is 56%, the largest of all countries (see the first, dark blue bar in Figure A.1a in the Online Appendix). This may be driven by politically motivated citations, which may be especially strong in power-oriented societies such as China (Jia et al., 2019). To understand whether this home citation share is 'abnormally' large, we follow Qiu et al. (2021) by using a 'dartboard approach' and consider a world without home bias, where one would expect the total citations in the world (where i and j index countries),

$$world\_cites := \sum_i \sum_j citations_{ij},$$

to be distributed across countries depending on the size of all citing and cited countries. Then, a benchmark for the citations of country i from country j can be determined by the expected citations of country i's articles by country j's articles based on the share of citations country i receives from the world, $\frac{\sum_j citations_{ij}}{world\_cites}$, times the size of the potential citing country based on its publication share, $\frac{pubs_j}{\sum_k pubs_k}$:

$$citations_{ij}^{benchmark} := world\_cites \times \frac{\sum_j citations_{ij}}{world\_cites} \times \frac{pubs_j}{\sum_k pubs_k} \tag{1}$$

Figure A.1a in the Online Appendix plots the benchmark share of home citations according to this expression, i.e., $\frac{citations_{ii}^{benchmark}}{\sum_j citations_{ij}}$, as a second, light blue bar.[11] China's benchmark share is also large, 18.5%, due to China's size in terms of publications, but what is even more striking is the gap between the benchmark share and the share of total citations accounted by home citations: its level is 37.7%, by far the largest of all countries, even though all of them exhibit a home bias of some sort.[12]

We adjust the abnormally large number of home citations by using a country-specific debiasing factor δ_h which we define as the number of benchmark home citations given by equation (1) divided by the actual home citations:

$$\delta_h := \frac{citations_{hh}^{benchmark}}{citations_{hh}^{actual}}$$

To get to our final quality measure, we therefore change the 'naive' measure of quality as given by an article's total citations, $citations_p = citations_{p,h} + citations_{p,US} + citations_{p,ROW}$, in two ways. First, we take out citations from the US, and second, we scale home citations by the country-specific debiasing factor δ_h:[13]

$$quality_p := \delta_h citations_{p,h} + citations_{p,ROW}$$

---

[11] These calculations are based on the full set of publications in Chemistry between 2000 and 2021. We do not drop any citations, and thus we assign a fraction of each citation to the citing and cited countries based on the share of the respective countries according to the addresses of affiliations listed on the article (fractional counts).

[12] This is consistent with the findings in Qiu et al. (2021) using different data. Furthermore, in Qiu et al. (2021) we show that the distribution of the home bias across countries is not sensitive to the specific way of defining the benchmark.

[13] The distribution of δ_h across countries is given in Figure A.1b of the online appendix. Note that our estimation results are not very sensitive to this adjustment. In Table B.2 of the online appendix we show that our results hold when we drop home citations completely (column 1); do not discount home citations at all (i.e., match on all citations outside US; column 3); or do not match on citations at all (column 4).

For the coarsened matching we divide this quality measure, which we label as the number of home-debiased citations, into six bins: $0-25^{th}$ percentile; $25-50^{th}$ percentile; $50-75^{th}$ percentile; $75-95^{th}$ percentile; $95-99^{th}$ percentile; and the top percentile.[14]

One may worry that citations from the rest of the world (ROW) are endogenous, if they are themselves affected by citations from the US. For example, if a Chinese article is cited less by US scientists, this may also lead to fewer citations by scientists in other countries. However, this spillover effect would bias us *against* finding a citation discount for Chinese articles: due to the relative underciting of Chinese articles, we would effectively compare a higher quality Chinese article with a lower-quality non-Chinese article, which should make it harder to estimate a negative citation effect for the Chinese article. A similar rationale would lead us to underestimate (in magnitudes) a Chinese citation discount if citations from ROW were strategic. For example, if Chinese articles are less likely to be authored by journal editors (as we will show in Table 4), they may receive fewer citations from ROW compared to articles from other countries, if PIs are more likely to cite editors in expectation of favorable treatment in the peer review process.

Additional matching variables. Besides the citation measure, at the article level, we match exactly on the journal, the publication year, and coarsely on the number of authors (4 groups: 1-3; 4-6; 7-9; 10 or more coauthors). At the researcher level, our list of matching covariates includes year of PhD receipt (in three-year bins). The union of all matching criteria defines a strata. Within each strata, articles are indistinguishable from the perspective of the CEM algorithm, and the matching is performed at the level of the strata.[15]

This procedure yields 6,905 treated articles written by 155 Chinese PIs, and 9,287 control articles written by 402 non-Chinese PIs. On average, there are 1.34 control articles per treated article.

Table 4 compares the characteristics of control and treated articles. The first four rows display the variables used for matching. By construction, the two sets of articles were published in the same year. Due to coarsened matching, the remaining variables are not identical across treatment and control group, but differences in both the mean and median are small and statistically insignificant. The control and treated articles have on average about 5 to 5.2 authors, and are written by investigators who graduated on average in 1994. Treated

---

[14] When creating these bins, we compute a separate empirical distribution of citations for each publication year.

[15] As there may be different numbers of treated and control articles in different strata, CEM assigns a weight to each matched article to adjust for strata size, and we use this weight in all regressions models.

articles received around 18 debiased non-US citations, only slightly fewer than control articles. In fact, Figure 2 indicates that the entire distribution of citations received from non-US countries is well balanced across treated and control articles.

Among the characteristics not used for matching, some significant differences subsist. For example, if anything, according to the total number of citations, Chinese articles should be of better 'quality' than the control articles, so this should make it less likely for us to find a China citation discount. In order to not shrink the size of the article sample any further, we do not add additional matching criteria; instead, our regression specifications will include these covariates as controls. For example, treated articles are less likely to have US coauthors and are less likely to be written by editorial authors, which may explain why Chinese articles are cited less by US authors; but we are able to include covariates in the specification that should alleviate this concern. The combination of matching and covariate inclusion results in comparisons that flexibly and plausibly hold "fertility" constant across Chinese and non-Chinese publications.

There is widespread skepticism regarding the quality of Chinese academic research (e.g., Xie and Freeman 2019). Research undertaken by Chinese scientists is sometimes alleged to consist of imitations or replications rather than original research, and other times assumed to be less well written (Boroush, 2020). However, the extant evidence is considerably more ambiguous than anecdotes would imply. Packalen (2019) uses Natural Language Processing techniques to analyze the novelty of research across nations, and concludes that Chinese research has in large part caught up to the world scientific frontier. Conversely, Bornmann et al. (2018) show that Chinese research are less relevant for elite publications, as measured by their prevalence—relative to China's publication share—in the references of highly-cited articles. Our prior is that such article characteristics would naturally be reflected in lower citations for Chinese articles outside of China and the US. As such, our research design indirectly addresses these concerns.

5.2 Definition of the risk set of citing articles

To test whether articles in the control or treatment group are cited differentially by US authors, we first need to determine which US articles are *at risk* of citing the articles in our sample, not just the articles that correspond to *actual* citations. Moreover, since we would like to evaluate how social or geographic proximity shapes the propensity to cite, it is crucial that participation in the risk set not be mechanically influenced by such factors. We therefore deem an article eligible to be part of the citation risk set if it is *topically* related to the article in our sample.

In order to specify topical relationships between articles, we rely on the fact that most of the articles in our sample (and most of the citations to these articles) appear in journals indexed by *PubMed* in addition to the *Web of Science*.[16] We then use the "Related Articles" function in *PubMed* to harvest journal articles that are intellectually proximate to the articles in our sample. This functionality is based on a topic-based content similarity model called *PubMed Related Citations Algorithm* or *PMRA* (Lin and Wilbur, 2007). This algorithm yields relatedness rankings and scores between any two articles based on the extent to which two articles are similar with respect to titles, abstracts, and keywords.[17] The *PMRA* algorithm is designed to estimate the conditional probability that a researcher would be interested in another article, given her interest in a focal article. For each article in our data, its citation risk set includes every PMRA neighbor whose authors work in US institutions and appeared after the focal article was published. Of the 43,979 US articles actually citing the 16,192 articles in the matched sample, only 6,272 (14.27%) correspond to related records in the sense of PMRA. Importantly, the risk set does not include actual citations that are PMRA-unrelated.[18]

The combined risk set of the 16,192 articles in the matched sample comprises 188,753 citable/potentially citing article pairs, with each article having on average 11.7 potentially citing articles in its risk set.

5.3 Model specification

We model the probability that article i is cited by each article $j \in J_i$, the citation risk set for article i, as a function of the characteristics of article i and article pair ij, using the following linear probability model:

---

[16] *PubMed* is an online resource from the National Library of Medicine that provides free and comprehensive access to the biomedical research literature, indexing more than 40,000 journals within the life sciences, including almost all the journals in which elite scientists routinely publish.

[17] To facilitate the harvesting of *PubMed*-related records on a large scale, we have developed an open-source software tool that queries *PubMed* and PMRA and stores the retrieved data in a MySQL database. The software is available for download at `http://www.stellman-greene.com/FindRelated/`. Prior research leveraging the intellectual linkages between articles generated by PMRA include Azoulay et al. (2015, 2019a), and Myers (2020). Appendix C in Azoulay et al. (2019a) describes the algorithm in detail. Importantly, that the PMRA algorithm does *not* take citations into account—it only takes as input titles words, abstract words, and keywords.

[18] In a robustness test in Table B.4 of the online appendix, we have verified that leaving these unrelated citations in the risk set does not alter our substantive conclusions.

$$\mathbb{1}_{(j \text{ cites } i)} = \beta_0 + \beta_1 China_i + \beta_2 X_{ij} + \beta_3 China_i \times X_{ij} + \varphi(i,j) + \varepsilon_{ij} \tag{2}$$

The dependent variable is an indicator variable that takes on value 1 if article j actually cites article i, and 0 otherwise. Our main regressor of interest, $China_i$, is an indicator variable for whether i's last author is Chinese, whereas X is a vector of control covariates, and $\varphi(i,j)$ corresponds to a large set of fixed effects for i and $i \times j$ characteristics. These include fixed effects for: (a) each strata defined in the coarsened exact matching algorithm (the interaction of our bins for journal × publication year × number of authors × debiased non-US citations × PhD degree year; yielding 3,847 bins); (b) investigator cumulative publication bins (13 bins), (c) investigator cumulative citations bins (13 bins), (d) investigator gender, (e) topic similarity rank bins,[19] (f) the interaction of i and j publication years (262 bins); (g) an indicator variable for the case when citing and cited articles were published in the same journal. We do not report coefficient estimates for these covariates, but they are always included.

We include additional controls X_{ij} to explore whether the China citation discount is driven by certain channels. Specifically, we explore whether the discount exists because of communication barriers between US and Chinese researchers, or because Chinese researchers have weaker networks with US scientists, or because Chinese researchers specialize in subfields that are less active in the US, or because Chinese investigators are not focused enough to disseminate their research, or perhaps because Chinese research is of lower *perceived* quality. Below we explain how we constructed the covariates whose inclusion in the regression model help us capture the impact of these channels or mechanisms.

Communication. A natural explanation for the China citation discount may be that there are language problems that restricts the communication between US and Chinese researchers and thus their awareness of published articles (even though we only consider articles in English journals); or that China is just far away which also reduces awareness of scientific output. We capture these potential channels by using two covariates: (a) an indicator variable for PI countries (as defined in section 3) that list English among their

---

[19] We group similarity ranks between articles into 26 bins, with finer bins for rankings below 150 (10 ranks per bin), and gradually create larger bins for higher ranks (151-170, 171-190, 191-210, 211-230, 231-260, 261-300, 301-350, 351-450, 451-650, 651-1000, above 1000).

official languages; and (b) the geographic distance between the city of the PI and the city of the affiliations of the US citing authors.[20]

Network. The dissemination of research may depend on PIs' access to potential citers via formal or informal networks. We include several covariates to capture the impact of this channel: (a) an indicator variable denoting whether PIs have obtained training (PhD or postdoc) in the US; (b) an indicator variable denoting whether the cited article's reprint or first author has a US affiliation;[21] (c) an indicator variable for whether any middle author has a US affiliation;[22] (d) an indicator variable that captures shared ethnicity between the PI and at least one author from article j;[23] (e) an indicator variable for the presence of a past coauthor of the PI on article j's authorship roster; (f) an indicator variable for the presence of a common author on the authorship rosters of articles i and j; and (g) an indicator variable for PIs that have written editorials, which proxies for editors and other influential scientists in the literature, who may be cited more for strategic reasons.

Geographic topical clustering. To the extent that researchers in certain countries concentrate in different subfields, it is important for the analysis to control for these country-level specialization patterns.[24] Rather than assigning each source article to a subfield arbitrarily, we rely on the PMRA tool described earlier and define the subfield of each source article as the set of its PMRA-neighbors, counting only the neighbors whose similarity score is above 0.5

---

[20] Specifically, we compute the log average geographical distance between the PI city of affiliation and all cities listed in the citing article j.

[21] Recall that only the cited article's last author is constrained to be a non-US researcher. First- and middle-authors in articles published by non-US PIs can be affiliated with US research institutions, and it is plausible that such coauthorships elevate the propensity of citation by other US-based researchers. In order to assign countries to specific authors, we need to link the address lines listed on articles to authors; this is possible for 80% of the articles in our database. For these articles we find that in most cases (95%), the first authors are linked to the first address record. Therefore, for the remaining 20% of articles for which we cannot accurately assign countries to specific authors, we use the first address line to define the first author's country.

[22] According to publication conventions in Chemistry, first or reprint authors have contributed significantly to the research undertaken.

[23] To identify the ethnic origin of US-based scholars, we map a scholar's last name to its ethnic origins based on the algorithm pioneered by Nguyen (2019). Her novel algorithm computes the probability that a last name corresponds to a particular ethnic origin based on the de-anonymized full population samples of US Censii between 1910 and 1940. A last name can be assigned to multiple origin countries with different probabilities, which are based on the relative frequencies of surname-origin pair appearing in the population samples of US Censii. We use the country with the maximum probability to identify the ethnic origin of each last name. More details regarding the name-ethnicity mapping can be consulted in Nguyen (2019).

[24] The existence of such patterns is not mere speculation on our part. For instance, Borjas and Doran (2015) document the persistence of Russian influence in certain mathematical subfields even after the dissolution of the Soviet Union.

and which appeared before the source article. Using these PMRA-derived subfields, we construct three subfield-level covariates: (i) the subfield's *home-research intensity* corresponds to the sum of the PMRA-relatedness scores for the articles in the subfield whose researchers are from the PI's country; (ii) the subfield's *foreign-research intensity* corresponds to the sum of the PMRA-relatedness scores for the articles in the subfield whose researchers are not from the PI's country and not from the US; and (iii) the subfield's *US-research intensity* corresponds to the sum of the PMRA-relatedness scores for the articles in the subfield whose researchers are from the US.

Investigator's intellectual focus. It is possible that investigators who concentrate their research in specific subfields receive higher recognition and thus more citations from the US, and researchers may be differently focused across countries. For this purpose we use our subfield definitions based on PMRA again and specify the three following measures: (i) *the subfield's importance for the investigator*; i.e., the number of articles for a given PI that belong to the subfield of the focal article divided by the total number of articles authored by the PI; (ii) *the investigator's importance for the subfield*, i.e., the number of articles of a given PI that belong to the focal article's subfield divided by the total number of articles in the subfield; (iii) *the investigator's research portfolio focus*, computed as an index to measure a PI's topical concentration across articles.[25]

Reputation. US researchers may be hesitant to cite articles if they appear in subfields with a reputation for questionable ethical standards. We construct an indicator variable which denotes whether a subfield is *retraction heavy*, i.e., whether there exists, among a cited article's PMRA neighbors, at least one article that has been either retracted, or has been the object of an 'expression of concern' or erratum.[26]

---

[25] Inspiring ourselves from Ellison and Glaeser (1997), we propose an index which 'normalizes' the PI's topical distribution of articles by the topical distribution of articles in the underlying population. For each PI and year, we consider the set of last-authored articles published by the PI in the previous five years, and assign each article to several research topics m based on the MeSH (Medical Subject Headings) keywords. Define N_{it} as the number of last-authored articles that PI i published in the five years before t; $Share_{imt}$ as the share of the MeSH terms of the N_{it} articles that are assigned to MeSH research topic m, and x_{mt} as the share of MeSH terms m in publications that appeared in year t. We define the Ellison-Glaser index as:

$$eg\_index_{it} := \frac{N_{it}}{N_{it}-1} \times \frac{\sum_m (Share_{imt} - x_{mt})^2}{1 - \sum_m x_{mt}^2} - \frac{1}{N_{it}-1} \tag{3}$$

[26] Because these events are rare, in this case we do not apply the 0.5 cutoff for these articles' relatedness score. In a robustness check, we found that imposing the cutoff weakens the precision of the corresponding coefficient estimate, but does not change its magnitude. Although we only count retractions of articles that

We cluster standard errors simultaneously at the level of the individual PI—to allow for arbitrary correlation of citation patterns across publications within each individual researcher—and the level of a strata—to allow for correlation of citation patterns across publications within a strata (Cameron and Miller, 2015).

6 Empirical Results

Table 5 reports the estimation results corresponding to equation (2). Column 1 only includes the Chinese investigator indicator variable, which is negative and significant, suggesting that Chinese articles face a lower probability of being cited in US research. Column 2 explores potential channels through which Chinese research may face this citation discount by progressively including into the specification the covariates specified in section 5.3.

The comparison between columns 2 and 1 reveals that while some of the control variables affect the likelihood of US citations, the magnitude of the Chinese PI effect appears impervious to their inclusion in the specification. For example, the communication controls do not affect the China discount, and also do not explain citation patterns: Articles from English-speaking countries do not receive more citations from the US; and PIs from cities farther away from the citing author's US city do not appear to receive fewer US citations.

Investigators with US training are cited more by US authors, so a US education probably increases the reach of PIs' US network (controlling for the quality of the research). Past as well as current co-authors are also more likely to cite articles, most likely because they are more aware of the focal article (as we already control for topical relatedness when we construct the risk set). In contrast, the presence of US coauthors, a common ethnicity by cited and citing authors, or being an editorial author does not in general increase the propensity of being cited by US PIs. Overall, differential network reach does not fully account for the citation discount that Chinese PIs experience on average.

The spatial clustering of research fields also has significant effects on citations from the US, but is not correlated with the China effect. For example, articles written in subfields that are intensely researched by non-US countries are cited less by the US. On the other hand, articles written in subfields which are strong in the US are in fact cited more by US articles. The intellectual focus of the PI also matters. articles that are written in subfields that are closely related to the other publications of the PI are associated with an increased

---

were published before the source, the retraction event can occur either before, in the same year, or after the publication of the source.

rate of US citations. The same is true for articles in subfields within which the PI is an important contributor globally. Articles belonging to retraction-heavy subfields are cited less themselves, but this effect is not statistically different from zero.

Across all specifications, we observe a statistically significant and negative "China effect": articles written by Chinese PIs receive significantly fewer citations from US scientists than articles written by non-Chinese PIs. The magnitude of the effect is empirically meaningful: Since the baseline probability of being cited by a US article is low in our sample (3.2%), the probability of a Chinese-authored article being cited is 28.1% lower than the baseline probability (based on the estimates from column 2, our baseline).

One may wonder whether the Chinese citation discount exists because the emergence of Chinese science is quite recent. In this case, one may expect the China discount to become smaller over time.[27] In Figure A.2 of the online appendix we estimate the Chinese discount separately by PI cited article publication year. There is no discernible pattern in the discount over time, but it is negative in almost all years, and statistically significant for many of the years. Overall, it does not seem that we can expect the Chinese citation discount to be a transitory phenomenon. In fact, one might expect an even larger Chinese citation discount since 2018, when the U.S. Department of Justice started the China Initiative which resulted in investigations of several hundred researchers that were collaborating with Chinese scientists (Jia et al., 2022). In fact, the point estimate in 2018 in Figure A.2 is negative, but we would need to wait several more years in order to study this effect systematically.

Another question is whether China's experience is unique, or whether other countries suffer from the same bias. In fact, the choice of China to define the treated group of articles is arbitrary. Would we find similar evidence of a discount if we chose to make researchers from other countries with a storied legacy in chemical research pivotal? In Figure 3 we replicate our analysis by making the articles from PIs located in other top Chemistry countries the treated group (Chinese PIs' articles are eligible to participate in the the control group). Among the six countries that have at least 1,500 articles in the matched sample, no other country experiences a significant citation discount, and the magnitude of the discount is also largest for China. Switzerland and Germany, two countries which are renowned for their important chemical industries, experience citation premia of 74% and 59%, respectively.

---

[27] For example, Iaria et al. (2022) show that the gender gap in citations has declined to zero over the course of the 20^{th} century.

6.1 What moderates the China citation discount?

So far we have established that Chinese elite chemists experience a citation discount from the US *on average*. We now turn to analyzing heterogeneous effects; asking whether some Chinese PIs can overcome the discount, whether the discount is less severe in some subfields, or whether some US researchers are less biased against Chinese research. We check this by allowing the China effect to vary across a number of characteristics that may be expected to matter more or less for Chinese research.[28]

For example, in column 3 of Table 5, we test whether strong networks of Chinese PIs, e.g., due to US training or US based coauthors, help Chinese articles overcome the citation discount. It turns out that the positive effect on US citations from column 2 is entirely driven by Chinese PIs, whereas the effect is insignificant for other countries. However, US education is not enough for Chinese researchers to overcome the US citation discount, it only reduces it by about half, but is still significantly different from zero.[29] We do not find significant effects from having US coauthors, neither for Chinese articles nor those from other countries, but it is notable that Chinese articles that have US first or reprint authors experience a further negative (though insignificant) effect on citations, while this is not the case for articles from other countries. One explanation for this could be that potential citers discount such instances of collaboration because they suspect the presence of the US author on the authorship roster to reflect scientifically "impure" motives, such as the need to curry favor with editors of leading journals.[30] We also examine whether US-based researchers who have ethnic roots in China help diffuse Chinese research to the US, as suggested by recent research (Xie and Freeman, 2020). This seems true, as we find a positive interaction effect on ethnicity, indicating that US researchers with a Chinese name do *not* cite Chinese articles less than articles from other countries.[31] This could also mean that Chinese researchers have access to ethnic-Chinese researchers in the US in their network, but not to other US researchers.

---

[28] We do not include interaction effects with all of the characteristics of column 2 of Table 5; we only included characteristics that we deemed ex ante to potentially matter differently for Chinese research. If we were to include interaction terms with the remaining characteristics, all but the interaction term with log(distance) would be insignificantly different from zero (not reported).

[29] The China effect for PIs with US training is -0.0065 (p-value 0.048).

[30] This interpretation is quite speculative. Perhaps the most infamous case of "ghost authorship"—a fraudulent article in the field of stem cell research—embroiled a South Korean team, not a Chinese one (Hwang et al., 2005).

[31] The China effect for PIs who are cited by US authors with Chinese ethnicity is the sum of the main estimate for China combined with the corresponding interaction effect. The magnitude of the combined effect in an imprecisely estimated -0.0046.

Column 4 checks whether the specialization of China in certain subfields has implications on being cited in US research. We do not find effects that are significantly different from zero. In column 5 we study whether Chinese PIs who are especially focused in their research overcome the China bias. We find that this is the case for Chinese PIs for whom the focal article's subfield is important, i.e., Chinese PIs that are publishing in their area of expertise. However, the China effect only disappears for the most focused PIs, i.e., those in the 99^{th} percentile of the subfield importance distribution.[32] It could be the case that US researchers are more aware of focused Chinese PIs, or that the research by these specialized Chinese PIs is taken more seriously.

Because of the relatively high frequency of retraction scandals that have afflicted Chinese scientific teams (Liao et al., 2018; Huang, 2017), we speculate that non-Chinese scientists could deem knowledge and ideas that originate in China to be less reliable than those originating in other countries, leading researchers to cite Chinese research less heavily even when it would appear equally fruitful based on other observable covariates. In column 6, we test this conjecture by interacting the Chinese PI indicator with a dummy that indicates the existence of retractions in the subfield of the focal article. We do find evidence of an additional citation discount imposed on Chinese articles that belong to subfields that are relatively more "retraction-heavy", but the corresponding estimate is not statistically significant, and the magnitude of the Chinese PI effect barely changes when controlling for this channel of perceived reputation.

Column 7 allows for all interaction effects to enter the specification simultaneously, with similar results. Overall, these specifications point towards an obdurate citation discount experienced by articles published by elite Chinese chemists, that can only be overcome in a handful of contingencies.

Next, we ask whether the magnitude of the discount is modulated by the underlying quality of the article, which we assess in two alternative ways in Figure 4. In Figure 4a we estimate heterogeneous effects by the quality of the journal of the cited article, as measured by its journal impact factor (JIF). Notice that since our PIs are star scientists, the majority of the articles we consider is published in the top 10% of journals. Nonetheless, we see that the bias is driven by the best among those journals, the top 5% of journals.[33] In Figure 4b

---

[32] The China effect for PIs in the 99^{th} percentile of the distribution of the importance of the subfield for the PI is the sum of the main estimate for China combined with the corresponding interaction effect. The magnitude of the combined effect in an imprecisely estimated 0.0047.

[33] The top 5% of Chemistry journals according to JIF are *Journal of the American Chemical Society*, *Angewandte Chemie International Edition*, *Nature Chemistry*, *Analytical Chemistry*, *Accounts of Chemi-*

we assess quality by the debiased citations it received from non-US sources. We interact the Chinese PI indicator with indicator variables for the position of the focal article in the citation distribution of all other articles with the same publication year. We create 11 percentile bins, allowing for more heterogeneity at the top of the distribution.[34] To allow for comparability of the estimates across percentile bins—which have a different baseline probability of being cited by US authors—we plot the coefficients measured in units of standard deviations. We find that the bias is especially pronounced at the middle of the distribution, i.e., those that lie between the 50^{th} and 70^{th} percentile of the impact distribution, but the effect is negative throughout.

One may also expect that the China citation discount would be dampened in the case of more senior PIs, as awareness of their research would increase the longer they had been active in the research community. In Table B.3 of the online appendix we proxy for seniority in different ways. Columns 1 and 2 proxy for seniority using the age of the PI. One may expect that the bias falls as PIs get older and better known; and this seems to be true, as the bias is not significantly different from zero for PIs who completed their PhD before 2014. However, columns 3-6 show a different pattern when we measure seniority based on the stock of publications or total citations a researcher received until one year before the publication of the focal article; in this case, the bias is stronger for the more accomplished PIs.

As an alternative way to control for citation dynamics, in Figure A.3 we restrict our citations to articles that were published within 3 or 5 years of the focal article. One may expect the citation discount to become stronger, the larger the time window we consider for citations to take place, if initial underciting amplifies over time. However, the pattern looks fairly stable as we increase the time window from 3 years to 5 years to all years in our data, but standard errors become smaller as we accrue observations in the estimation sample.

6.2 Chinese address versus Chinese name

So far, we have identified a stable China discount across a variety of specifications that is unique to China. Is this a reflection of animus towards Chinese researchers, rather than reduced awareness or a reduced integration into the US research community? This question is hard to answer, but in Table 6 we test whether elite PIs that have Chinese names but are working outside of China (but not in the US) experience the same bias. Interestingly, the

---

cal Research, *Chemistry*, *Chemical Science*, *Chemical Society Reviews*, *The Journal of Physical Chemistry Letters*, *Nano Letters*, *Nanoscale*, *Nature Protocols*, *Small*, *Nucleic Acids Research* and *ChemSusChem*.

[34] More specifically, we create percentile bins at the following cutoffs: 10^{th}, 20^{th}, 30^{th}, 40^{th}, 50^{th}, 60^{th}, 70^{th}, 80^{th}, 90^{th}, and 95^{th} percentile.

bias is specific to PIs working in China, and not present for PIs with Chinese names that undertake their research in other countries.[35]

7 Leveraging article citations from patents

More than a fundamental scientific discipline, the field of Chemistry also forms the basis for technological advances in industry, including the biopharmaceutical sector (Adams, 1990). In addition, patenting is a common way for firms to appropriate the returns from innovation in this domain (Cohen et al., 2000), since at least the emergence of the periodic table of elements in the late 19^{th} century (Moser, 2012).

Recently, it has become possible to track citations made by patents to the open scientific literature at scale (Marx and Fuegi, 2020; Roach and Cohen, 2013), thus providing a lens on understanding how advances in basic science percolate in industry R&D (Azoulay et al., 2019b). We leverage this novel source of data by studying the extent to which US patent inventors rely on scientific research in Chemistry originating from China versus other countries.

The research design parallels the one used for the analysis of article-to-article citations. To begin, we focus on the full set of chemistry articles published between 2000 and 2018 (as in Table 3), changing the outcome variables from article citations to US patent citations with all-American inventor teams. We find that the China discount is also present in patent citations. In the raw data, Chinese articles receive 0.083 cites from US patents on average, while non-Chinese articles receive 0.255 cites. Table 7 provides the results of Poisson regressions. Column 1 shows that a Chinese article received on average 70% fewer citations by US patents compared to a article from other non-US countries. The magnitude of the estimate barely changes when we add fixed effects for the number of authors (column 2), and is roughly halved when adding publication year effects in column 3. Column 4 includes journal fixed effects in the specification, which further reduces the magnitude of the China citation discount to 25%. Focusing on within-journal variation can be thought of as a crude way to hold the quality of the underlying articles constant. Below, we go further by performing a careful matching on article quality, as well as researchers' patenting activities, based on the data set of articles written by our elite chemistry PIs.

---

[35] PIs with Chinese names that are working in non-Chinese research institutions may not be suitable members of the control group. There are 26 such PIs in our data. In Table B.5, we check whether our results change when dropping these PIs from the estimation sample. Reassuringly, we find that our results are unaffected.

We perform a careful matching on article quality as in the analysis for article-to-article citations, and go further by matching also on a researchers' patenting activities.

Scientists who are active inventors themselves may have networks that reach beyond academia into industry, thereby heightening both awareness and relevance of their investigations in the eyes of R&D intensive firms (which account for the lion's share of all inventive activity, cf. Azoulay et al. 2012). Therefore, we modify the coarsened exact matching strategy used in Section 5 and enrich it with measures of PI's patenting activities, which we track by linking investigators' names to the names of inventors on USPTO patents. This mapping process is challenging, as we must guard against mistakenly assigning a patent to a scientist when the invention was actually performed by a namesake. In order to accurately attribute patents to our set of elite investigators, we conduct extensive manual checks which take into account data fields such as institutional affiliation, city, country, and research interest that overlap between the CV and the patent information. We identify 5,562 patents which list one of our 751 elite PIs as an inventor: 48.1% of Chinese investigators and 61.7% of other non-US investigators have been granted at least one patent by the USPTO.

Next, we gather a list of patent citations to our PIs' articles, which we extract from the dataset constructed by Marx and Fuegi (2020). This process results in a list of 47,831 patent-to-article citations for the 78,541 PI articles. Each article receives on average 0.609 patent citations overall, but only 0.310 patent citations from all-US inventor teams. Importantly, 87% of the articles in the sample are not cited by any US patent. As we did for the analysis of article-to-article citations, we restrict the patent-to-article citations to those from inventions with US inventors only.

A key limitation of the analysis presented below is that we do not have at our disposal a set of patents that are at risk of citing each article—the *PubMed* Related Citations Algorithm identifies topically-close articles, but there is no equivalent algorithm to identify topically-close patents: Every observation in the data corresponds to an actual citation. As a result, we aggregate the number of patent citations up to the article level and estimate Poisson models to analyze the determinants of the count of US patents for each article via Quasi-Maximum Likelihood, with robust standard errors clustered at the level of the PI country.

Column 1a and column 1b of Table 8 display results based on a coarsened exact matching approach very similar to that used earlier in Table 5. Column 1a implies that an article written by Chinese investigators receives 30% fewer citations from US patents, relative to an article written by other non-US investigators.[36] After controlling for characteristics of articles and investigators, as shown in column 1b, Chinese publications' citation discount in US patents slightly increases to 35%.

Columns 2a and 2b repeat the analysis, but with a more demanding matching approach which incorporates the PI's patenting history. Namely, to pair each Chinese article with an article by a non-Chinese (and non-US) PI, we require that the scientists have the same patent inventor status (an indicator variable) in addition to a match on the baseline set of covariates. Columns 3a and 3b report estimates based on an even finer match, by breaking down the patenting stock up to the focal article's year of publication into four groups: no patent application, exactly one patent application, exactly two patent applications, and three or more patent applications. The China citation discount remains stable: after controlling for article and researcher characteristics, column 2b shows that an article authored by a Chinese PI receives 41% fewer citations from US patents, and in column 3b, the discount is 42%.

The last pair of columns in Table 8 further imposes a coarse match on home-debiased patent-to-article citations from outside the United States based on the same method used to correct for the home bias in article citations.[37] Specifically, we match on articles' home-debiased patent citations from outside the US, split into four bins: 0, 1, 2, or 3+ citations. The results are presented in columns 4a and 4b. Once again, Chinese publications exhibit a stable citation discount in patents compared to publications from other countries, though the effect is only statistically significant at the 10% level.

We conclude, based on this slightly more aggregated analysis, that US industrial firms (which account for the bulk of patent citations) tend to build less on scientific research originating from Chinese labs, in a fashion similar to our earlier finding that US academics (which account for the bulk of article citations) appear to discount Chinese research, relative to research originating from other countries.

---

[36] Since $\exp(-0.362) - 1 = -0.30$. See the last row of Table 8 for equivalent % changes.

[37] To generate the patent-to-article citation benchmark, we follow equation (1) but replace the article-to-article citations $citations_{ij}$ with patent-to-article citations from country i's articles to country j's patents, and the publication share $\frac{pubs_j}{\sum_k pubs_k}$ with the country j's granted patent share in the USPTO data.

8 Conclusion

The inclusion of Chinese scientists in the global “Republic of Science” has gathered pace over the past two decades. An increasing body of evidence points to a gradual bridging of the gap that long existed between the impact of Chinese published scientific output and that of frontier countries (Xie and Freeman, 2019). Observers note—with a mix of awe and trepidation—that Chinese scientists are about to overtake US scientists in at least one domain: Artificial Intelligence (O’Meara, 2019).

Whereas the “quality view” stresses the broader shoulders provided by Chinese researchers for follow-on scientific developments (wherever they come from), the “spillover view” emphasizes the localized nature of much of the citations accruing to Chinese articles. Our study purposefully sidesteps this debate to shed light on the propensity to cite research emanating from Chinese scientists *holding quality constant*, by pairing Chinese and non-Chinese articles well matched on attributes that plausibly capture the scientific “fertility” of each publication. Focusing on elite researchers in a single domain—Chemistry—we uncover the existence of a sizable citation discount for Chinese articles, relative to non-Chinese articles. What explains the relative underciting of Chinese science by US scientists?

One possibility is that in spite of our best efforts, systematic differences in citation potential subsist between treated and control articles, even after carefully matching on journal and citations received from non-US sources. Another possibility is that our results are driven by differences in *perceived* quality between Chinese and non-Chinese articles. In this case it might reflect animus directed at Chinese scientists, but this hypothesis does not sit well with the evidence that the citation discount is not apparent for researchers with Chinese names doing research outside China. The discount might also reflect perceptions of lower reliability for Chinese-produced knowledge (Liao et al., 2018). These perceptions might arise due to the number of well-publicized cases of scientific misconduct in China (Huang, 2017). Azoulay et al. (2015) find that areas of science tainted by retraction scandals experience an exodus from scientists in that field which likely corresponds to an “overcorrection.” However, we do not think that accurate or inaccurate beliefs regarding the lower reliability of science produced in China is the full explanation for the results. While the bias is stronger against Chinese PIs publishing in subfields where retractions occurred, this effect is not statistically significant.

Another possibility is that US scientists are simply less aware of Chinese research, perhaps because Chinese scientists, even if they belong to the elite, have less access to the networks that provide broad exposure to research findings. This may explain the discount to some extent, as it is partly overcome by returnees who completed their scientific training in the US, as it is not present for an author with a Chinese name on the potentially citing article, and as it is reduced at least for Chinese PIs that are very specialized in their subfield.

Is the China citation discount likely to be a transitory phenomenon? The absence of a bias against older Chinese PIs would argue in favor of this view, yet overall the Chinese citation discount has proven fairly stable over the past two decades. If awareness and networking are the explanations, current US-China tensions, as well as the disruption of scientific travel induced by the COVID-19 pandemic, might further solidify the lower awareness of foreign citers vis-à-vis research produced in China (Jia et al., 2022).

References

Adams, James D. 1990. "Fundamental Stocks of Knowledge and Productivity Growth." *Journal of Political Economy* **98**(4):673-702.

Agnew, Neville. 1997. *Conservation of Ancient Sites on the Silk Road: Proceedings of an International Conference on the Conservation of Grotto Sites.* Los Angeles, CA: Getty Conservation Institute.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S. Graff Zivin. 2019a. "Does Science Advance One Funeral at a Time?" *American Economic Review* **109**(8):2889-2920.

Azoulay, Pierre, Joshua S. Graff Zivin, Danielle Li, and Bhaven N. Sampat. 2019b. "Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules." *The Review of Economic Studies* **86**(1):117-152.

Azoulay, Pierre, Jeffrey L. Furman, Joshua L. Krieger, and Fiona Murray. 2015. "Retractions." *The Review of Economics and Statistics* **97**(5):1118-1136.

Azoulay, Pierre, Joshua S. Graff Zivin, and Bhaven N. Sampat. 2012. "The Diffusion of Scientific Knowledge Across Time and Space: Evidence From Professional Transitions for the Superstars of Medicine." Chapter 2 in *The Rate Direction of Inventive Activity Revisited*, edited by Joshua Lerner and Scott Stern, pp. 107-155. Chicago, IL: University of Chicago Press.

Bai, Chunli. 2000. "Development and Prospect of Chinese Chemistry." *University Chemistry* **15**:1-10 (in Chinese).

Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Porro. 2009. "`cem`: Coarsened Exact Matching in Stata." *The Stata Journal* **9**(4):524-546.

Borjas, George J. and Kirk B. Doran. 2015. "Cognitive Mobility: Labor Market Responses to Supply Shocks in the Space of Ideas." *Journal of Labor Economics* **33**(S1):S109–S145.

Bornmann, Lutz, Caroline Wagner, and Loet Leydesdorff. 2018. "The Geography of References in Elite Articles: Which Countries Contribute to the Archives of Knowledge?" *PLoS One* **13**(3): e0194805.

Boroush, Mark. 2020. "Research and Development: US Trends and International Comparisons." in *Science and Engineering Indicators 2020,* NSB-2020-3. Arlington, VA: National Science Foundation.

Cameron, A. Colin and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-robust Inference." *Journal of Human Resources* **50**(2):317-372.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2000. "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)." NBER Working Paper #7552.

Ellison, Glenn and Edward L. Glaeser. 1997. "Geographic Concentration in US Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* **105**(5):889-927.

Gaulé, Patrick and Mario Piacentini. 2013. "Chinese Graduate Students and US Scientific Productivity." *The Review of Economics and Statistics* **95**(2):698-701.

Huang, Echo. 2017. “China Publishes More Science Research with Fabricated Peer-Review than Everyone Else Put Together.” *Quartz*, May 8. Accessed at `https://qz.com/978037/china-publishes-more-science-research-with-fabricated-peer-review-than-everyone-else-put-together`

Hwang, Woo Suk, Sung Il Roh, Byeong Chun Lee, Sung Keun Kang, Dae Kee Kwon, Sue Kim, Sun Jong Kim, Sun Woo Park, Hee Sun Kwon, Chang Kyu Lee, et al. 2005. “Patient-Specific Embryonic Stem Cells Derived From Human SCNT Blastocysts.” *Science* **308**(5729):1777-1783.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger. 2022. “Gender Gaps in Academia: Global Evidence over the Twentieth Century.” CEPR Discussion Paper #17422.

Jia, Ruixue, Huihua Nie, and Wei Xiao. 2019. “Power and Publications in Chinese Academia.” *Journal of Comparative Economics* **47**(4):792-805.

Jia, Ruixue, Margaret E. Roberts, Ye Wang, and Eddie Yang. 2022. “The Impact of U.S.-China Tensions on U.S. Science.” NBER Working Paper #29941.

Leicester, Henry M. 1971. *The Historical Background of Chemistry.* New York: Dover Publications.

Li, Chiao-Ping. 1948. *The Chemical Arts of Old China.* Easton: Journal of Chemical Education.

Liao, Qing-Jiao, Yuan-Yuan Zhang, Yu-Chen Fan, Ming-Hua Zheng, Yu Bai, Guy D. Eslick, Xing-Xiang He, Shi-Bing Zhang, Harry Hua-Xiang Xia, and Hua He. 2018. “Perceptions of Chinese Biomedical Researchers Towards Academic Misconduct: A Comparison Between 2015 and 2010.” *Science and Engineering Ethics* **24**(2):629-645.

Lin, Jimmy and W. John Wilbur. 2007. “PubMed Related Articles: A Probabilistic Topic-based Model for Content Similarity.” *BMC Bioinformatics* **8(1)**:423.

Marx, Matt and Aaron Fuegi. 2020. “Reliance on Science: Worldwide Front-page Patent Citations to Scientific Articles.” *Strategic Management Journal* **41**(9):1572-1594.

Moser, Petra. 2012. “Innovation Without Patents: Evidence from World’s Fairs.” *The Journal of Law and Economics* **55**(1):43-74.

Myers, Kyle. 2020. “The Elasticity of Science.” *American Economic Journal: Applied Economics* **12**(4):103-34.

Needham, Joseph, Ping-Yü Ho, Gwei-djen Lu, and Ling Wang. 1986. *Military Technology: The Gunpowder Epic.* Cambridge, UK: Cambridge University Press.

Nguyen, Kieu-Trang. 2019. “Trust and Innovation Within the Firm: Evidence from Matched CEO-Firm Data.” Working Paper, Northwestern University.

O’Meara, Sarah. 2019. “Will China Lead the World in AI by 2030?” *Nature* **572**(7770):427-428.

Packalen, Mikko. 2019. “Edge factors: Scientific Frontier Positions of Nations.” *Scientometrics* **118**(3):787-808.

Qiu, Shumin, Claudia Steinwender, and Pierre Azoulay. 2021. “Home Bias: China’s Trade in Science.” Working Paper, MIT.

Roach, Michael and Wesley M. Cohen. 2013. “Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research.” *Management Science* **59**(2):504-525.

Tollefson, Jeff. 2018. “China Declared World’s Largest Producer of Scientific Articles.” *Nature* **553**(7689):390.

Torvik, Vetle I. and Neil R. Smalheiser. 2009. “Author Name Disambiguation in MEDLINE.” *ACM Transactions on Knowledge Discovery from Data* **3**(3):11.

Tsuen-Hsuin, Tsien. 1985. “Chemistry and Chemical Technology: Paper and Printing.” Volume 5.1 in *Science and Civilization in China*, edited by Joseph Needham, pp. 132-184. Cambridge, UK: Cambridge University Press.

Xie, Qingnan and Richard B. Freeman. 2019. “Bigger Than You Thought: China’s Contribution To Scientific Publications and Its Impact on the Global Economy.” *China & World Economy* 27(1):1-27.

Xie, Qingnan and Richard B. Freeman. 2020. “The Contribution of Chinese Diaspora Researchers to Scientific Publications and China’s ‘Great Leap Forward’ in Global Science.” NBER Working Paper #27169.

Figures

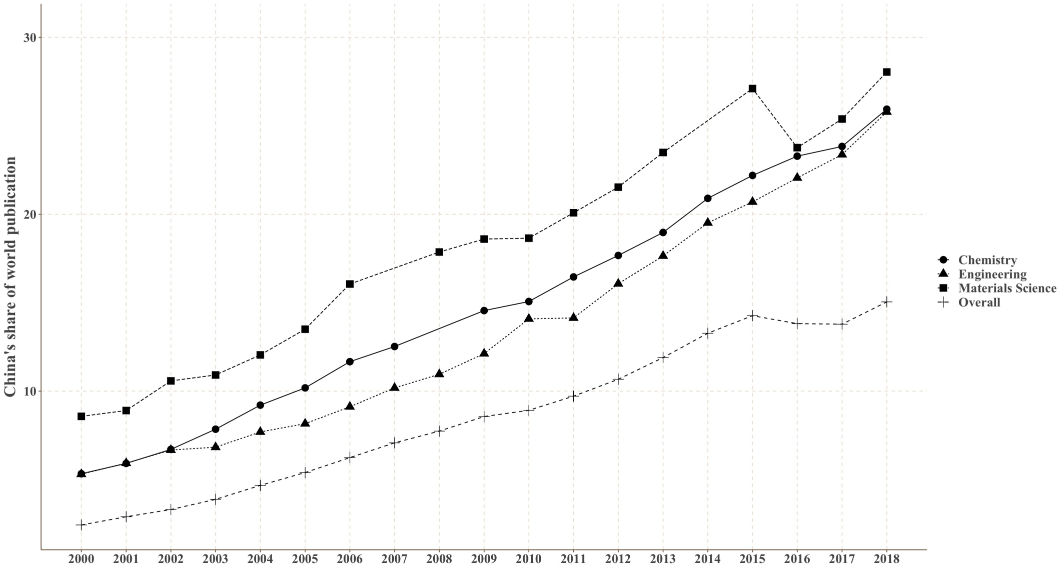


Figure 1: China's share of world publications, 2000-2018

Note: The share of publication is computed based on the share of Chinese addresses in English-language research articles in Web of Science, 2000-2018.

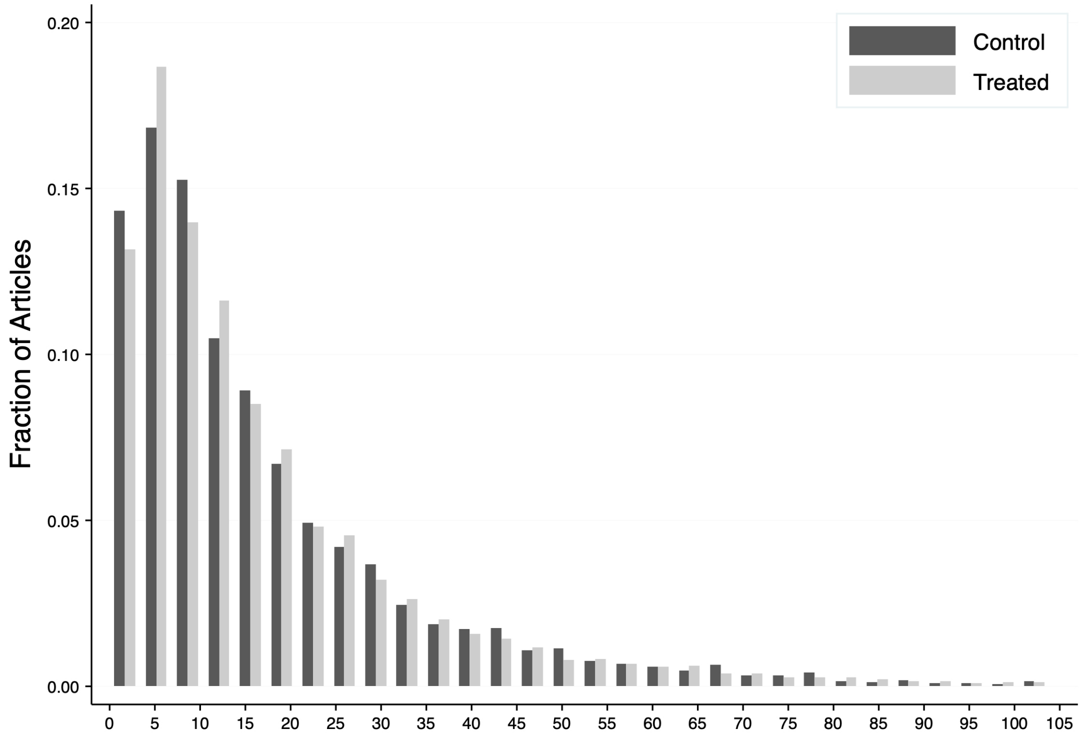


Figure 2: Distribution of non-US citations for control and treated articles

Note: The histogram for the number of home-debiased citations, excluding US citations is displayed above. The histogram excludes publications with 104 or more citations (approximately 1% of the sample).

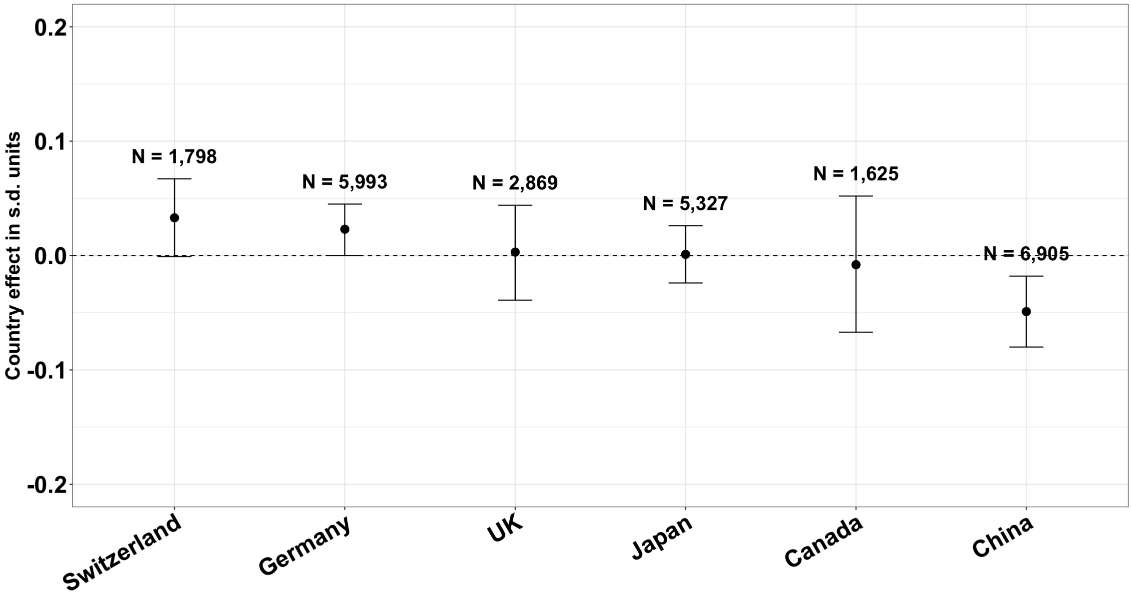
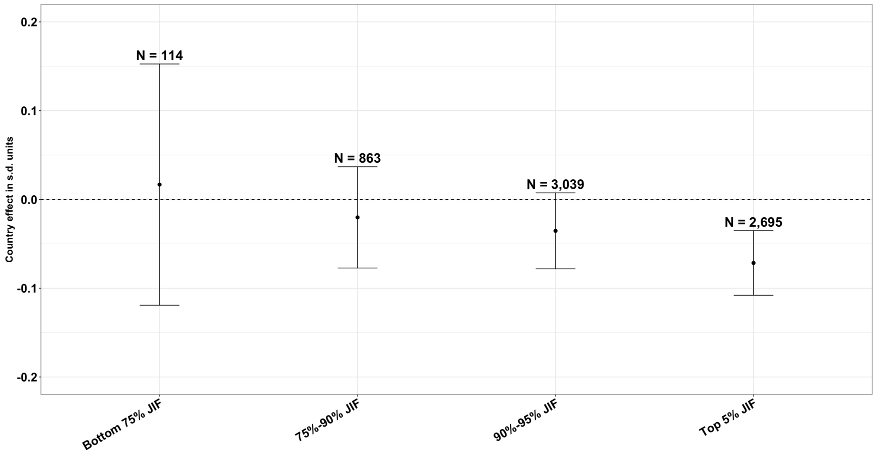
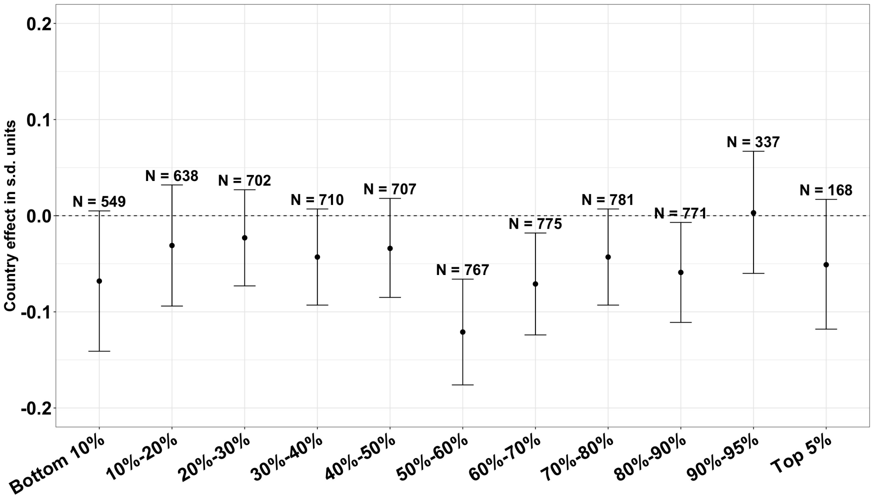


Figure 3: Heterogeneous effects of Chinese PIs on US citations, by source country

Note: We replace China with Switzerland, Germany, Canada, UK, Japan and Canada, respectively, to generate new treated and control groups, and estimate the country effect for each treated country with the same specification as column 2 in Table 5. The dark dots in the above plots correspond to country effect in standard deviation units for each treated country. The 95 percent confidence interval (the corresponding standard errors are two-way clustered at the investigator and matching strata levels) around these estimates is plotted with vertical lines. The number of treated articles for each country is indicated above the corresponding coefficient estimate.



(a) by journal impact factor (JIF)



(b) by home debiased citations outside US

Figure 4: Heterogeneous effect of Chinese PIs on US citations, by 'quality' of the cited article

Notes: The dark dots in the above plots correspond to coefficient estimates stemming from a Linear Probability Model in which the dependent variable is an indicator variable that equals 1 if the related article cited the source article, and 0 otherwise. In subfigure (a), the covariates of interest are 4 interaction terms between the China indicator variable and indicator variables for five quantiles of the distribution of source articles' Journal Impact Factor (JIF). The figure only shows source articles in Chemistry journals, though an interaction between the China effect and an indicator variable for non-chemistry journals is included in the corresponding specification. In subfigure (b), the covariates of interest are twelve interaction terms between the China indicator variable and indicator variables for various quantiles of the distribution of home-debiased non-US citations received. In both figures, the corresponding specification also includes all the covariates included in column 2 of Table 5. The 95 percent confidence interval (the corresponding standard errors are two-way clustered at the investigator and matching strata levels) around these estimates is plotted with vertical lines. The number of treated articles for each subgroup is indicated above the corresponding coefficient estimate.

Tables

Table 1: China's highly cited researchers (HCRs) and China's publications across fields

Field	No. of China's HCRs in field	China's share of world HCRs in field (%)	China's share in world publications in field (%)
Chemistry	211	19.27	14.96
Materials Science	205	26.12	19.99
Engineering	164	18.94	14.78
Computer Science	59	9.83	13.13
Physics	57	7.33	14.78
Mathematics	49	10.06	15.93
Geosciences	45	5.77	10.97
Molecular Biology/Biochemistry/Genetics	29	1.36	10.00
Plant/Animal Science	20	2.03	9.79
Agricultural Sciences	18	2.62	9.34
Pharmacology/Toxicology	10	1.44	10.42
Environment/Ecology	9	1.18	10.34
Neuroscience/Behavior	9	1.09	4.88
Microbiology	5	0.88	6.65
Immunology	5	0.83	6.27
Clinical Medicine	2	0.10	5.33
Psychiatry/Psychology	1	0.16	2.24

Notes: (1) Highly Cited Researchers (HCRs) are selected based on their production of multiple highly cited articles that rank in the top 1% by citations in a field and year (in the Web of Science database). (2) We count the number of HCRs of each country without dropping duplicates (i.e., the same person on the HCR list of different years is counted repeatedly); for researchers who are affiliated to more than one institution, we defined their affiliation (country) based on their primary institution in the year when the HCR report was issued. (3) The share is computed based on English-language research articles published in each field between 2000 and 2015. Articles are attributed to countries on the basis of the share of institutional addresses located in the country, relative to the total number of institutional addresses.

Table 2: China's research in Chemistry compared to other countries

Country/Region	Chemistry			All fields		
	Share of Articles	No. of HCRs	Share of HCRs	Share of Articles	No. of HCRs	Share of HCRs
United States	19.31	471	43.01	25.17	8,306	46.48
China	14.96	211	19.27	9.20	1,104	6.18
Japan	7.66	30	2.74	6.32	420	2.35
Germany	5.55	77	7.03	5.06	1,023	5.73
India	4.68	4	0.37	2.95	27	0.15
United Kingdom	4.20	31	2.83	5.70	1,701	9.52
France	3.94	27	2.47	3.57	474	2.65
South Korea	3.28	33	3.01	2.83	163	0.91
Italy	2.96	3	0.27	3.32	274	1.53
Spain	2.95	26	2.37	2.55	309	1.73
Canada	2.47	20	1.83	3.34	451	2.52
Taiwan	1.49	2	0.18	1.74	86	0.48
Australia	1.48	22	1.83	2.44	581	3.25
Iran	1.37	1	0.09	1.09	46	0.26
Switzerland	1.15	33	3.01	1.08	420	2.35
Netherlands	1.06	7	0.64	1.75	465	2.60
Sweden	0.97	1	0.09	1.23	163	0.91
Belgium	0.74	3	0.27	0.88	197	1.10
Czech	0.67	7	0.64	0.49	26	0.15
Israel	0.59	10	0.91	0.80	58	0.33
Singapore	0.56	19	1.74	0.53	164	0.92
Denmark	0.53	7	0.64	0.68	163	0.91
Hong Kong	0.43	12	1.10	0.57	144	0.81
South Africa	0.29	3	0.27	0.45	32	0.18
Ireland	0.26	5	0.46	0.33	79	0.44
Saudi Arabia	0.25	32	2.92	0.23	272	1.52
Rest of World	16.22	0	0.00	15.71	722	4.04

Notes: (1) We list 26 countries and a residual "rest of the world" category that jointly include all HCRs in Chemistry, ranked by the share of Chemistry articles they produce. (2) The share of Chemistry articles is computed based on English-language original research articles in Chemistry during the period 2000-2015. Articles are attributed to countries on the basis of the share of institutional addresses located in the country, relative to the total number of institutional addresses. (3) The figures correspond to *Highly Cited Researchers* reports during the 2014-2018 period, and excludes the social sciences and business categories when tallying the number of HCRs in a country. (4) We count the number of HCRs of each country without dropping duplicates (i.e., the same person on the HCR list of different years is counted repeatedly); for researchers who are affiliated to more than one institution, we defined their affiliation (country) based on their primary institution in the year when the *Highly Cited Researchers* report was issued.

Table 3: Effect of Chinese investigatorship on the number of US citations

	(1)	(2)	(3)
Chinese investigator	-0.662**	-0.416**	-0.269**
	(0.010)	(0.097)	(0.052)
articles	658,621	658,621	658,471
Publication year FE		yes	yes
Journal FE			yes
Pseudo R^2	0.024	0.089	0.261
% increase	-48%	-34%	-24%

Notes: The dependent variable is the number of US citations, i.e., citations from articles with only US-based authors. All regressions include fixed effects for the number of authors. The sample includes all articles in the field of Chemistry between 2000 and 2018, provided their authorship team hails from a single country (articles with geographically-mixed authorship teams are excluded). Coefficients derive from a Poisson specification estimated via quasi–maximum likelihood. Robust standard errors in parentheses are clustered by the country of the cited article. $^{\dagger}p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 4: Summary statistics of control and treated articles

	Control articles (N= 9,287)					Treated articles (N= 6,905)					Difference	
	Mean	Median	Std. Dev.	Min	Max	Mean	Median	Std. Dev.	Min	Max	Mean	Median
Characteristics used for matching:												
Source Article Pubyear	2011.879	2012	3.872	2000	2018	2011.879	2012	3.872	2000	2018	0	0
Source Article Number of Authors	5.034	5	1.975	1	15	5.174	5	1.926	1	15	-0.140	0
Investigator PhD Degree Year	1993.447	1994	6.277	1958	2008	1993.660	1994	6.212	1952	2007	-0.213	0
Number of Debiased Citations Outside US	18.568	11.603	24.103	0	625	17.926	11.645	21.653	0	598	0.642	-0.042
Characteristics not used for matching:												
No. of Total Citations	24.177	16	29.489	1	726	33.191	21	40.570	0	1,125	-9.014**	-5*
No. of Citations Outside US	20.661	13	25.585	0	638	30.187	19	37.014	0	1,017	-9.526**	-6*
US First/Reprinted Cited Author	0.018	0	0.133	0	1	0.006	0	0.077	0	1	0.012**	0
US Cited Author in Other Positions	0.037	0	0.189	0	1	0.016	0	0.127	0	1	0.021**	0
Retraction-heavy Subfield	0.028	0	0.164	0	1	0.043	0	0.204	0	1	-0.015**	0
Subfield Foreign Research Intensity	17.542	14.104	16.456	0	170.860	14.427	11.631	13.167	0	190.219	3.115**	2.473*
Subfield Home Research Intensity	1.639	0.757	2.452	0	38.926	11.278	7.993	12.402	0	158.569	-9.639**	-7.236*
Subfield USA Research Intensity	4.402	3.053	4.927	0	64.233	4.440	3.036	5.186	0	79.131	-0.038	0.017
Importance of Investigator for the Subfield	0.061	0.038	0.076	0	1	0.050	0.029	0.062	0	0.542	0.011**	0.009*
Importance of Subfield for Investigator	0.124	0.052	0.329	0	10	0.131	0.042	0.359	0	8.500	-0.007	0.01*
Ellison/Glaeser Index of Scholarly Focus	0.017	0.014	0.015	-0.026	0.270	0.015	0.013	0.012	-0.020	0.142	0.002**	0.001*
Female Investigator	0.038	0	0.191	0	1	0.046	0	0.209	0	1	-0.008*	0
Investigator with US Training	0.611	1	0.488	0	1	0.476	0	0.499	0	1	0.135**	1
Investigator is Editorial Author on Elite Journals	0.032	0	0.176	0	1	0.010	0	0.102	0	1	0.022**	0
Investigator from English-speaking Country	0.342	0	0.474	0	1	0	0	0	0	0	0.342**	0
Investigator of Chinese Names Working outside China	0.088	0	0.284	0	1	0	0	0	0	0	0.088**	0
Investigator Publication Stock	82.698	70	62.321	0	738	91.051	73	72.270	0	496	-8.353**	-3*
Investigator Citation Stock	1,305.696	729	1,638.410	0	13,082	1,163.350	643	1,435.532	0	11,243	142.346**	86*

Notes: The stock of publications and citations for each investigator are assessed at the end of the year prior the year of publication for each source article.

Table 5: Estimating the China location discount (or premium) on the rate of US citations [Linear Probability Model]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Chinese Investigator	-0.008**	-0.009**	-0.013**	-0.009**	-0.010**	-0.008**	-0.014**
	(0.002)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)
Communication							
Investigator from English-speaking Country		-0.001	-0.000	-0.000	-0.001	-0.001	-0.000
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Log(Avg. Distance)		0.003	0.003	0.003	0.003	0.003	0.003
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Network							
Investigator with US Training		0.005*	0.002	0.005*	0.005*	0.005*	0.002
		(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
US First/Reprinted Cited Author		0.011	0.016	0.011	0.011	0.011	0.016
		(0.010)	(0.012)	(0.010)	(0.010)	(0.010)	(0.012)
US Cited Author in Other Positions		0.006	0.005	0.005	0.006	0.006	0.005
		(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)
Citation from Same Ethnicity		0.003	-0.001	0.003	0.003	0.003	-0.001
		(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Citing Coauthor is Investigator's Past Collaborator		0.050**	0.050**	0.050**	0.050**	0.050**	0.049**
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Common Coauthor		0.166**	0.166**	0.166**	0.166**	0.166**	0.166**
		(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Investigator is an Editorial Author		-0.003	-0.001	-0.003	-0.003	-0.003	-0.001
		(0.011)	(0.012)	(0.011)	(0.011)	(0.011)	(0.012)
Geographic Topical Clustering							
Subfield Home Research Intensity		0.009	0.010	0.029	0.010	0.010	0.034
		(0.009)	(0.009)	(0.048)	(0.009)	(0.009)	(0.048)
Subfield Foreign Research Intensity		-0.022**	-0.022**	-0.028**	-0.022**	-0.022**	-0.029**
		(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
Subfield USA Research Intensity		0.044*	0.045*	0.057†	0.045*	0.044*	0.061*
		(0.020)	(0.020)	(0.030)	(0.020)	(0.020)	(0.030)
Investigator's Intellectual Focus							
Importance of Subfield for Investigator		0.004*	0.004†	0.004†	0.001	0.004*	0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Importance of Investigator for the Subfield		0.033*	0.032*	0.033*	0.039*	0.033*	0.038*
		(0.016)	(0.016)	(0.016)	(0.019)	(0.016)	(0.019)
Ellison/Glaeser Index of Scholarly Focus		0.062	0.055	0.062	0.043	0.062	0.039
		(0.053)	(0.053)	(0.053)	(0.064)	(0.053)	(0.064)
Reputation							
Retraction-heavy Subfield		-0.003	-0.003	-0.003	-0.003	-0.000	0.000
		(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Interactions with Network							
Chinese Investigator × Investigator with US Training			0.006†				0.007†
			(0.004)				(0.004)
Chinese Investigator × US First/Reprinted Cited Author			-0.028				-0.025
			(0.019)				(0.019)
Chinese Investigator × US Cited Author in Other Positions			0.004				0.003
			(0.010)				(0.010)
Chinese Investigator × Citation from Same Ethnicity			0.008*				0.008*
			(0.004)				(0.004)
Chinese Investigator × Investigator is an Editorial Author			-0.005				-0.005
			(0.020)				(0.020)
Interactions with Geographic Topical Clustering							
Chinese Investigator × Subfield Home Research Intensity				-0.026			-0.029
				(0.050)			(0.050)
Chinese Investigator × Subfield Foreign Research Intensity				0.018			0.018
				(0.016)			(0.016)
Chinese Investigator × Subfield USA Research Intensity				-0.035			-0.040
				(0.036)			(0.035)
Interactions with Investigator's Intellectual Focus							
Chinese Investigator × Importance of Subfield for Investigator					0.006*		0.006*
					(0.003)		(0.003)
Chinese Investigator × Importance of Investigator for the Subfield					-0.015		-0.017
					(0.034)		(0.034)
Chinese Investigator × Ellison/Glaeser Index of Scholarly Focus					0.052		0.042
					(0.104)		(0.103)
Interactions with Reputation							
Chinese Investigator × Retraction-heavy Subfield						-0.004	-0.006
						(0.007)	(0.007)
Mean of Dependent Variable	0.032	0.032	0.032	0.032	0.032	0.032	0.032
s.d. of Dependent Variable	0.177	0.177	0.177	0.177	0.177	0.177	0.177
China effect in s.d. units	-0.043	-0.049	-0.072	-0.049	-0.057	-0.048	-0.077
Adjusted R^2	0.079	0.084	0.084	0.084	0.084	0.084	0.084
No. of Investigators	557	557	557	557	557	557	557
No. of Cited Articles	16,192	16,192	16,192	16,192	16,192	16,192	16,192
No. of Citing Articles	71,409	71,409	71,409	71,409	71,409	71,409	71,409
No. of Citing/Cited Article Pairs	188,753	188,753	188,753	188,753	188,753	188,753	188,753

Notes: The dependent variable is an indicator variable that equals 1 if the related article cites the PI's article, and 0 otherwise. All regressions include fixed effects for rank bins of each citing article j with respect to its topic similarity to article i; fixed effects for the interaction of citing and cited article publication year; fixed effects for each CEM strata; fixed effects for the investigator's highest degree year, a investigator gender indicator variable, and an indicator if citing and cited articles are published in the same journal (coefficients not reported). Standard errors in parentheses are two-way clustered at the investigator and strata level. $^{\dagger}p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 6: China discount for researchers based in China vs. researchers with a Chinese name

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Chinese Investigator	-0.008**	-0.008**	-0.012**	-0.008*	-0.010*	-0.008**	-0.013**
	(0.002)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)
Chinese Names Working outside China	-0.001	0.004	0.004	0.004	0.004	0.004	0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Communication							
Investigator from English-speaking Country		-0.001	-0.001	-0.001	-0.001	-0.001	-0.000
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Log(Avg. Distance)		0.003	0.002	0.003	0.003	0.003	0.002
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Network							
Investigator with US Training		0.005*	0.002	0.005*	0.005*	0.005*	0.002
		(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
US First/Reprinted Cited Author		0.011	0.016	0.011	0.011	0.011	0.016
		(0.010)	(0.012)	(0.010)	(0.010)	(0.010)	(0.012)
US Cited Author in Other Positions		0.006	0.005	0.005	0.006	0.006	0.005
		(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)
Citation from Same Ethnicity		0.003	-0.001	0.003	0.003	0.003	-0.001
		(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Citing Coauthor is Investigator's Past Collaborator		0.050**	0.050**	0.050**	0.050**	0.050**	0.049**
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Common Coauthor		0.166**	0.166**	0.166**	0.166**	0.166**	0.166**
		(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Investigator is an Editorial Author		-0.002	-0.001	-0.002	-0.003	-0.002	-0.001
		(0.011)	(0.012)	(0.011)	(0.011)	(0.011)	(0.012)
Geographic Topical Clustering							
Subfield Home Research Intensity		0.010	0.010	0.036	0.010	0.010	0.041
		(0.009)	(0.009)	(0.047)	(0.009)	(0.009)	(0.047)
Subfield Foreign Research Intensity		-0.022**	-0.023**	-0.029**	-0.023**	-0.022**	-0.030**
		(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
Subfield USA Research Intensity		0.044*	0.045*	0.058†	0.045*	0.044*	0.062*
		(0.020)	(0.020)	(0.030)	(0.020)	(0.020)	(0.030)
Investigator's Intellectual Focus							
Importance of Subfield for Investigator		0.004*	0.004†	0.004†	0.001	0.004*	0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Importance of Investigator for the Subfield		0.033*	0.032*	0.033*	0.039*	0.034*	0.038*
		(0.016)	(0.016)	(0.016)	(0.019)	(0.016)	(0.019)
Ellison/Glaeser Index of Scholarly Focus		0.064	0.056	0.064	0.045	0.063	0.041
		(0.053)	(0.053)	(0.053)	(0.064)	(0.053)	(0.064)
Reputation							
Retraction-heavy Subfield		-0.003	-0.003	-0.003	-0.003	-0.000	0.000
		(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Interactions with Network							
Chinese Investigator × Investigator with US Training			0.006†				0.007†
			(0.004)				(0.004)
Chinese Investigator × US First/Reprinted Cited Author			-0.028				-0.026
			(0.019)				(0.019)
Chinese Investigator × US Cited Author in Other Positions			0.003				0.003
			(0.010)				(0.010)
Chinese Investigator × Citation from Same Ethnicity			0.008*				0.008*
			(0.004)				(0.004)
Chinese Investigator × Investigator is an Editorial Author			-0.006				-0.005
			(0.020)				(0.020)
Interactions with Geographic Topical Clustering							
Chinese Investigator × Subfield Home Research Intensity				-0.033			-0.035
				(0.050)			(0.050)
Chinese Investigator × Subfield Foreign Research Intensity				0.019			0.019
				(0.016)			(0.016)
Chinese Investigator × Subfield USA Research Intensity				-0.036			-0.041
				(0.036)			(0.035)
Interactions with Investigator's Intellectual Focus							
Chinese Investigator × Importance of Subfield for Investigator					0.006*		0.006*
					(0.003)		(0.003)
Chinese Investigator × Importance of Investigator for the Subfield					-0.014		-0.016
					(0.034)		(0.034)
Chinese Investigator × Ellison/Glaeser Index of Scholarly Focus					0.051		0.041
					(0.104)		(0.103)
Interactions with Reputation							
Chinese Investigator × Retraction-heavy Subfield						-0.004	-0.006
						(0.007)	(0.007)
Mean of Dependent Variable	0.032	0.032	0.032	0.032	0.032	0.032	0.032
s.d. of Dependent Variable	0.177	0.177	0.177	0.177	0.177	0.177	0.177
China effect in s.d. units	-0.044	-0.047	-0.071	-0.047	-0.056	-0.046	-0.075
Adjusted R^2	0.079	0.084	0.084	0.084	0.084	0.084	0.084
No. of Investigators	557	557	557	557	557	557	557
No. of Cited Articles	16,192	16,192	16,192	16,192	16,192	16,192	16,192
No. of Citing Articles	71,409	71,409	71,409	71,409	71,409	71,409	71,409
No. of Citing/Cited Article Pairs	188,753	188,753	188,753	188,753	188,753	188,753	188,753

Notes: The dependent variable is an indicator variable that equals 1 if the related article cites the PI's article, and 0 otherwise. All regressions include fixed effects for rank bins of each citing article j with respect to its topic similarity to article i; fixed effects for the interaction of citing and cited article publication year; fixed effects for each CEM strata; fixed effects for the investigator's highest degree year and a investigator gender indicator variable (coefficients not reported). Standard errors in parentheses are two-way clustered at the investigator and strata level. $^{\dagger}p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 7: Effect of Chinese investigatorship on the number of US Patent citations [Poisson Model]

	(1)	(2)	(3)	(4)
Chinese investigator	-1.124*	-1.171**	-0.506**	-0.284**
	(0.103)	(0.105)	(0.078)	(0.056)
Articles	658,621	658,621	658,621	651,872
Number of author FEs		yes	yes	yes
Publication year FE			yes	yes
Journal FE				yes
Pseudo R^2	0.018	0.020	0.166	0.255
% increase	-68%	-70%	-40%	-25%

Notes: The dependent variable is the number of US patent citations, i.e., patent citations from all-US inventor teams. The sample includes all articles in the field of Chemistry between 2000 and 2018, provided their authorship team hails from a single country (articles with geographically-mixed authorship teams are excluded). Coefficients derive from a Poisson specification estimated via quasi–maximum likelihood. Robust standard errors in parentheses are clustered by the country of the cited article. $^{\dagger}p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

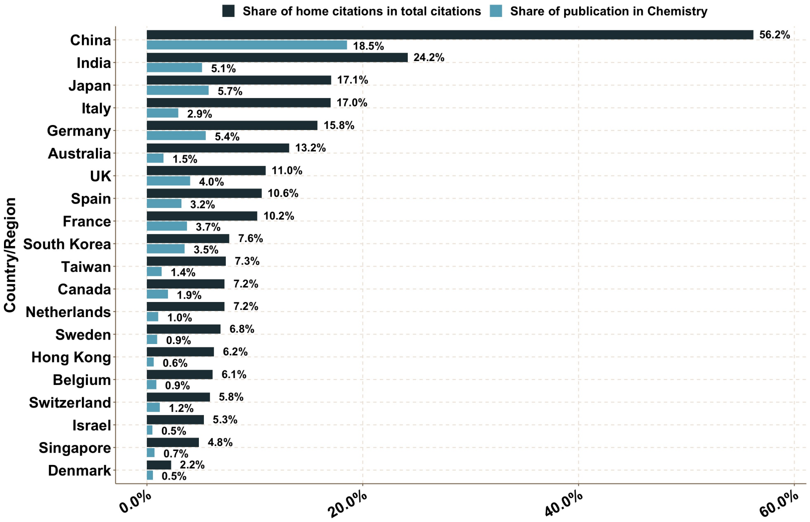
Table 8: Effect of Chinese investigatorship on the number of US Patent citations [Poisson Model]

Matching on:	article-to-article citations (baseline)		baseline + patent inventor status		baseline + patent stock categories		baseline + patent inventor status + patent-to-article citations	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Chinese Investigator	-0.362*	-0.430*	-0.367*	-0.533*	-0.369†	-0.552*	-0.482*	-0.425†
	(0.159)	(0.197)	(0.182)	(0.236)	(0.197)	(0.254)	(0.216)	(0.224)
Investigator from English-speaking Country		0.030		-0.048		-0.150		0.031
		(0.225)		(0.270)		(0.284)		(0.286)
Investigator with US Training		-0.190		-0.156		-0.196		0.139
		(0.172)		(0.197)		(0.206)		(0.196)
US Citded Author(s)		-0.318		-0.637*		-0.908**		-1.043**
		(0.265)		(0.323)		(0.300)		(0.303)
Subfield Home Research Intensity		-0.090		0.867		0.646		0.553
		(0.809)		(0.915)		(0.969)		(0.895)
Subfield Foreign Research Intensity		-0.791		-1.049		-0.492		0.675
		(0.673)		(1.049)		(1.046)		(0.848)
Subfield USA Research Intensity		1.814		2.370		1.377		-0.936
		(1.322)		(2.003)		(2.138)		(1.981)
Importance of Subfield for Investigator		-0.418†		-0.363		-0.332		-0.093
		(0.239)		(0.229)		(0.227)		(0.211)
Importance of Investigator for the Subfield		-2.249†		-1.718		-1.387		-0.254
		(1.346)		(1.401)		(1.423)		(1.316)
Ellison/Glaeser Index of Scholarly Focus		-4.774		-5.811		-4.080		-1.503
		(3.364)		(4.408)		(4.397)		(5.002)
Retraction-heavy Subfield		0.034		-0.264		-0.432		-1.585**
		(0.404)		(0.479)		(0.584)		(0.484)
Investigator Publication Stock (log)		0.127		0.023		-0.075		-0.552*
		(0.155)		(0.184)		(0.208)		(0.237)
Investigator Citation Stock (log)		-0.046		0.001		0.031		0.310*
		(0.108)		(0.123)		(0.140)		(0.143)
Pseudo R^2	0.178	0.188	0.207	0.217	0.203	0.213	0.189	0.201
Cited Articles	16,192	16,192	16,089	16,089	13,448	13,448	14,915	14,915
Investigators	557	557	545	545	527	527	542	542
% increase	-30%	-35%	-31%	-41%	-31%	-42%	-38%	-35%

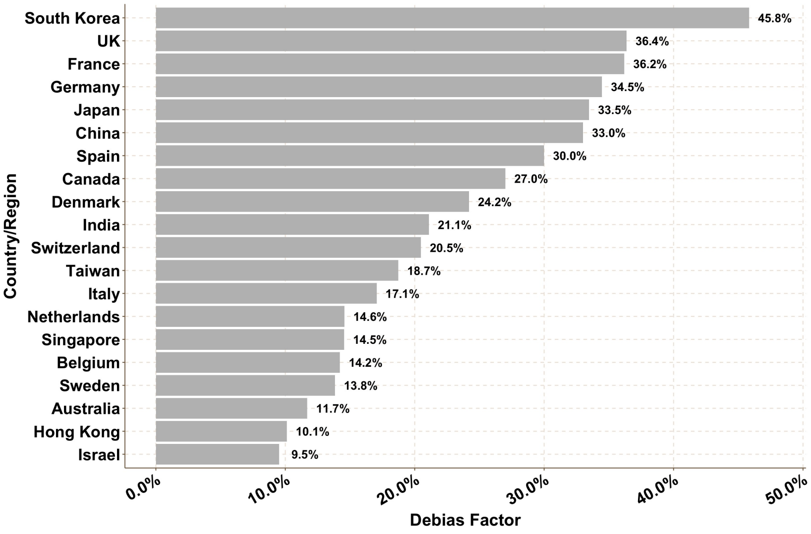
Notes: The dependent variable is the cumulative number of US patent citations received by a published article, i.e., patent citations from all-US inventor teams. Table 4, Table B.6, Table B.7 and Table B.8 respectively provide descriptive statistics for control and treated articles that form the estimation samples used in Columns 1a and 1b, columns 2a and 2b, columns 3a and 3b, and columns 4a and 4b. All specifications include fixed effects for the cited article's publication year, journal, number of authors, as well as indicator variables for the investigator's highest degree year and investigator gender (coefficients not reported). Coefficients derive from a Poisson specification estimated via quasi–maximum likelihood. Robust standard errors in parentheses are clustered at the level of the investigator. $^{\dagger}p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

ONLINE APPENDIX

A Appendix Figures



(a) Actual home citation share versus benchmark share



(b) Debias factors by country/region

Figure A.1: Home bias in citations

Note: These figures include the top 20 countries according to citations; in the analysis all countries are included.

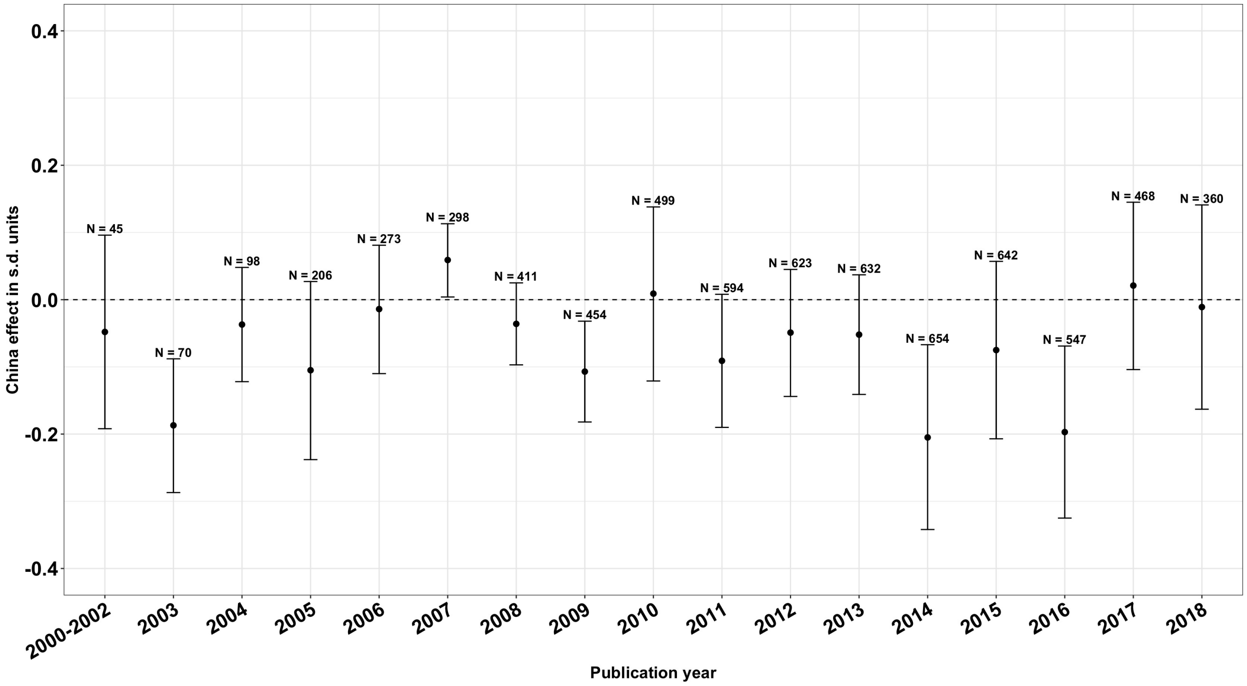


Figure A.2: The China effect over time

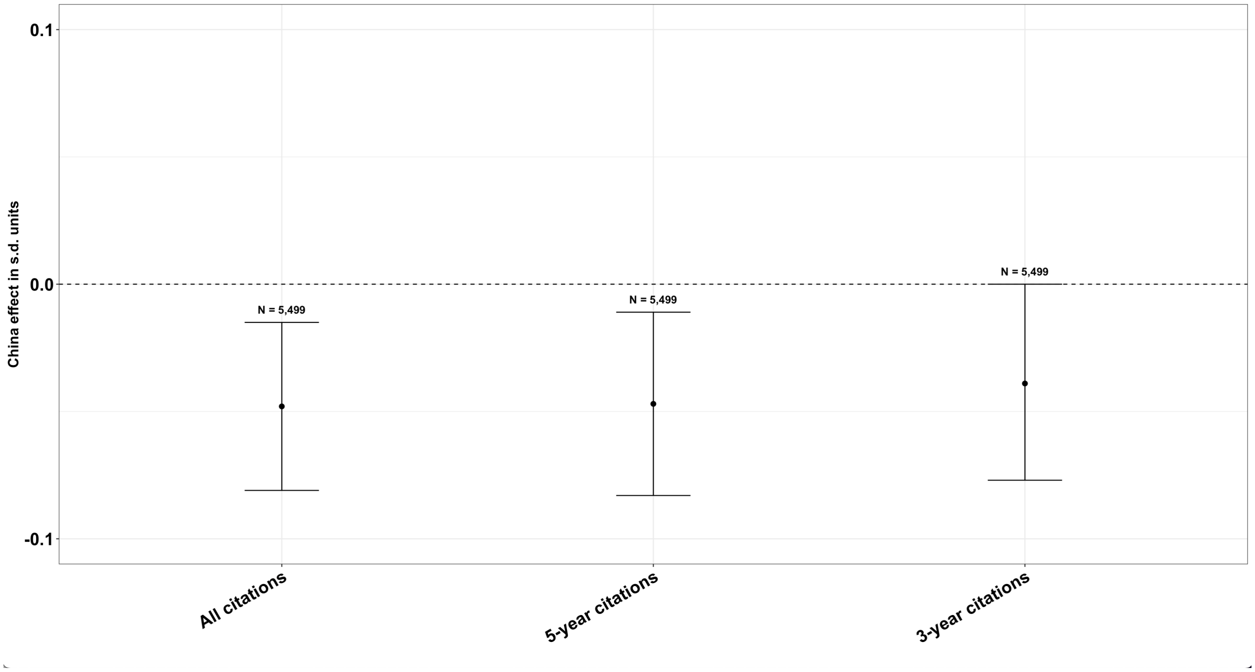


Figure A.3: Restricting citations to different time windows after publication

Notes: We estimate the same specification as in column (2) of Table 5. We restrict the sample to cited articles that were published between 2000 and 2016, and consider either citations from all publication after the publication of the focal article, or from publications that occur within 5 or 3 years after the publication of the focal article, respectively.

B Appendix Tables

Table B.1: Elite journal list

Journal Name	2020 JIF	No. of Cites from 2020, Received by Articles 2011-2020	JIF without Self Cites	Total Articles 2018-2019
Nature Materials	43.841	112,429	43.542	194
Nature Nanotechnology	39.213	75,845	38.84	143
Nature Chemistry	24.427	41,139	24.235	150
Journal of the American Chemical Society	15.419	609,264	14.394	2,442
Angewandte Chemie-International Edition	15.336	410,750	13.98	3,599
Nature Chemical Biology	15.04	27,429	14.81	171
ACS Central Science	14.553	11,097	14.38	197
Nano Letters	11.189	177,909	10.776	1,148
Chemical Science	9.825	65,945	9.359	1,344
Chemistry of Materials	9.811	124,153	9.311	1,004
Journal of Medicinal Chemistry	7.446	85,946	6.629	889
Analytical Chemistry	6.986	156,738	6.087	1,992
Chemical Communications	6.222	213,293	5.879	3,020
Organic Letters	6.005	107,262	5.203	1,843
Macromolecules	5.985	111,371	5.158	1,055
Chemistry-A European Journal	5.236	110,572	4.867	2,075
Inorganic Chemistry	5.165	103,059	4.62	1,854
Journal of Biological Chemistry	5.157	397,474	4.969	1,295
Dalton Transactions	4.39	80,783	4.004	1,698
Journal of Organic Chemistry	4.354	101,397	3.94	1,501
Langmuir	3.882	125,608	3.523	1,608
Organometallics	3.876	36,815	3.509	494
Journal of Chemical Physics	3.488	231,490	2.739	2,008
Journal of Biochemistry	3.387	9,099	3.288	106
Analytical Biochemistry	3.365	42,956	3.268	303
Journal of Physical Chemistry B	2.991	111,523	2.68	1,130
Bioorganic & Medicinal Chemistry Letters	2.823	42,591	2.702	638
Journal of Physical Chemistry A	2.781	63,916	2.469	1,060
Tetrahedron	2.457	45,774	2.356	452
Tetrahedron Letters	2.415	60,350	2.277	771
Chemical Physics Letters	2.328	50,930	2.127	993

Notes: Journals are ranked by Impact Factor according to Journal Citation Report 2020.

Table B.2: Matching based on different quality measures

	Match on Citations Excluding Home	Match on Debiased Citations Outside US	Match on All Citations Outside US	Do not Match on Citations
Chinese Investigator	-0.006*	-0.009**	-0.012**	-0.009**
	(0.003)	(0.003)	(0.003)	(0.002)
Mean of Dependent Variable	0.031	0.032	0.035	0.035
S.d. of Dependent Variable	0.173	0.177	0.184	0.184
China effect in s.d. units	-0.034	-0.049	-0.065	-0.049
Adjusted R^2	0.079	0.084	0.087	0.069
No. of Investigators	560	557	553	567
No. of Cited Articles	15,502	16,192	16,103	27,065
No. of Citing Articles	70,819	71,409	71,903	100,819
No. of Citing/Cited Article Pairs	180,947	188,753	192,754	343,864

Notes: The dependent variable is an indicator variable that equals 1 if the related article cites the source article, and 0 otherwise. Fixed effects included as in the baseline, i.e., column 2 of Table 5. Standard errors in parentheses are two-way clustered at the investigator and strata level. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table B.3: Heterogeneity by investigator seniority

	PhD Year		Publication Stock		Citation Stock	
	after 2014	before 2014	below median	above median	below median	above median
Chinese Investigator	-0.012**	-0.005	-0.005†	-0.015**	-0.003	-0.016**
	(0.005)	(0.003)	(0.003)	(0.006)	(0.003)	(0.006)
Mean of Dependent Variable	0.035	0.030	0.027	0.042	0.026	0.042
s.d. of Dependent Variable	0.183	0.171	0.163	0.201	0.159	0.201
China effect in s.d. units	-0.067	-0.029	-0.032	-0.074	-0.022	-0.078
Adjusted R^2	0.087	0.081	0.075	0.116	0.066	0.111
No. of Investigators	238	319	527	488	541	498
No. of Cited Articles	7,843	8,349	8,281	7,683	7,757	8,259
No. of Citing Articles	42,023	51,612	54,428	35,301	53,811	35,259
No. of Citing/Cited Article Pairs	82,629	106,121	127,994	60,521	119,719	68,851

Notes: The dependent variable is an indicator variable that equals 1 if the related article cites the PI's article, and 0 otherwise. All regressions include fixed effects for rank bins of each citing article j with respect to its topic similarity to article i; fixed effects for the interaction of citing and cited article publication year; fixed effects for each CEM strata; fixed effects for the investigator's highest degree year, a investigator gender indicator variable, and an indicator if citing and cited articles are published in the same journal (coefficients not reported). Standard errors in parentheses are two-way clustered at the investigator and strata level. $^{\dagger}p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table B.4: Include unrelated actual citations in risk set

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Chinese Investigator	-0.006**	-0.006**	-0.010**	-0.006*	-0.007**	-0.006**	-0.010**
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Communication							
Investigator from English-speaking Country		-0.001	-0.000	-0.001	-0.001	-0.001	0.000
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Log (Avg. Distance)		0.002	0.001	0.002	0.002	0.001	0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Network							
Investigator with US Training		0.005**	0.003	0.005**	0.004**	0.005**	0.003
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
US First/Reprinted Cited Author		0.009	0.012	0.010	0.009	0.010	0.012
		(0.008)	(0.010)	(0.008)	(0.008)	(0.008)	(0.009)
US Cited Author in Other Positions		0.003	0.002	0.003	0.004	0.003	0.002
		(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
Citation from Same Ethnicity		0.000	-0.003†	0.000	0.000	0.000	-0.003†
		(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Citing Author is Investigator's Past Collaborator		0.039**	0.039**	0.039**	0.039**	0.039**	0.039**
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Common Coauthor		0.095**	0.095**	0.095**	0.095**	0.095**	0.095**
		(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Investigator is an Editorial Author		-0.001	0.003	-0.001	-0.001	-0.001	0.003
		(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)
Geographic Topical Clustering							
Subfield Home Research Intensity		0.008	0.009	0.030	0.009	0.009	0.034
		(0.007)	(0.007)	(0.037)	(0.007)	(0.007)	(0.037)
Subfield Foreign Research Intensity		-0.022**	-0.022**	-0.026**	-0.022**	-0.022**	-0.028**
		(0.006)	(0.006)	(0.008)	(0.006)	(0.006)	(0.008)
Subfield USA Research Intensity		0.036*	0.037*	0.046†	0.036*	0.036*	0.050†
		(0.017)	(0.017)	(0.027)	(0.017)	(0.017)	(0.026)
Investigator's Intellectual Focus							
Importance of Subfield for Investigator		0.004*	0.003*	0.003*	0.001	0.004*	0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Importance of Investigator for the Subfield		0.009	0.009	0.009	0.010	0.009	0.010
		(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.009)
Ellison/Glaeser Index of Scholarly Focus		0.059	0.057	0.059	0.054	0.059	0.053
		(0.043)	(0.043)	(0.043)	(0.052)	(0.043)	(0.052)
Reputation							
Retraction-heavy Subfield		-0.003	-0.003	-0.003	-0.003	-0.000	-0.000
		(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Interactions with Network							
Chinese Investigator × Investigator with US Training			0.004				0.004
			(0.003)				(0.003)
Chinese Investigator × US First/Reprinted Cited Author			-0.016				-0.015
			(0.016)				(0.016)
Chinese Investigator × US Cited Author in Other Positions			0.005				0.005
			(0.007)				(0.007)
Chinese Investigator × Citation from Same Ethnicity			0.005**				0.006**
			(0.002)				(0.002)
Chinese Investigator ×Investigator is an Editorial Author			-0.016				-0.015
			(0.015)				(0.015)
Interactions with Geographic Topical Clustering							
Chinese Investigator × Subfield Home Research Intensity				-0.025			-0.027
				(0.039)			(0.039)
Chinese Investigator × Subfield Foreign Research Intensity				0.012			0.013
				(0.013)			(0.013)
Chinese Investigator × Subfield USA Research Intensity				-0.026			-0.031
				(0.030)			(0.030)
Interactions with Investigator's Intellectual Focus							
Chinese Investigator × Importance of Subfield for Investigator					0.004		0.004
					(0.003)		(0.003)
Chinese Investigator × Importance of Investigator for the Subfield					-0.000		-0.003
					(0.019)		(0.019)
Chinese Investigator × Ellison/Glaeser Index of Scholarly Focus					0.015		0.010
					(0.084)		(0.083)
Interactions with Reputation							
Chinese Investigator × Retraction-heavy Subfield						-0.005	-0.006
						(0.005)	(0.005)
Mean of Dependent Variable	0.185	0.185	0.185	0.185	0.185	0.185	0.185
s.d. of Dependent Variable	0.388	0.388	0.388	0.388	0.388	0.388	0.388
China effect in s.d. units	-0.014	-0.016	-0.025	-0.015	-0.019	-0.016	-0.026
Adjusted R^2	0.837	0.838	0.838	0.838	0.838	0.838	0.838
No. of Investigators	557	557	557	557	557	557	557
No. of Cited Articles	16,192	16,192	16,192	16,192	16,192	16,192	16,192
No. of Citing Articles	80,088	80,088	80,088	80,088	80,088	80,088	80,088
No. of Citing/Cited Pairs	226,456	226,456	226,456	226,456	226,456	226,456	226,456

Notes: The dependent variable is an indicator variable that equals 1 if the related article cites the PI's article, and 0 otherwise. All regressions include fixed effects for rank bins of each citing article j with respect to its topic similarity to article i; fixed effects for the interaction of citing and cited article publication year; fixed effects for each CEM strata; fixed effects for the investigator's highest degree year, a investigator gender indicator variable, and an indicator if citing and cited articles are published in the same journal (coefficients not reported). Standard errors in parentheses are two-way clustered at the investigator and strata level. $^{\dagger}p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table B.5: Removing investigators with Chinese names that work outside China

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Chinese Investigator	-0.008**	-0.007*	-0.012**	-0.007*	-0.009*	-0.006*	-0.014**
	(0.002)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)
Communication							
Investigator from English-speaking Country		-0.001	-0.001	-0.000	-0.001	-0.001	-0.000
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Log (Avg. Distance)		-0.000	-0.001	-0.000	-0.000	-0.000	-0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Network							
Investigator with US Training		0.006**	0.003	0.006**	0.006**	0.006**	0.002
		(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
US First/Reprinted Cited Author		0.009	0.014	0.009	0.009	0.009	0.014
		(0.011)	(0.013)	(0.011)	(0.011)	(0.011)	(0.012)
US Cited Author in Other Positions		0.004	0.003	0.004	0.005	0.004	0.004
		(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)
Citation from Same Ethnicity		0.003	-0.001	0.002	0.003	0.003	-0.001
		(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Citing Coauthor is Investigator's Past Collaborator		0.052**	0.052**	0.052**	0.052**	0.052**	0.052**
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Common Coauthor		0.178**	0.178**	0.178**	0.178**	0.178**	0.178**
		(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Investigator is an Editorial Author		-0.004	-0.001	-0.004	-0.004	-0.004	-0.002
		(0.010)	(0.012)	(0.010)	(0.010)	(0.010)	(0.013)
Geographic Topical Clustering							
Subfield Home Research Intensity		0.005	0.006	0.061	0.006	0.007	0.065
		(0.009)	(0.009)	(0.048)	(0.009)	(0.009)	(0.048)
Subfield Foreign Research Intensity		-0.022**	-0.023**	-0.031**	-0.023**	-0.023**	-0.033**
		(0.008)	(0.008)	(0.011)	(0.008)	(0.008)	(0.011)
Subfield USA Research Intensity		0.046*	0.047*	0.052	0.047*	0.045*	0.057†
		(0.021)	(0.021)	(0.032)	(0.021)	(0.021)	(0.032)
Investigator's Intellectual Focus							
Importance of Subfield for Investigator		0.004*	0.004*	0.004*	0.001	0.005*	0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Importance of Investigator for the Subfield		0.034*	0.033*	0.034*	0.036†	0.034*	0.035†
		(0.016)	(0.016)	(0.016)	(0.019)	(0.016)	(0.019)
Ellison/Glaeser Index of Scholarly Focus		0.028	0.020	0.027	-0.004	0.028	-0.006
		(0.054)	(0.055)	(0.054)	(0.065)	(0.054)	(0.065)
Reputation							
Retraction-heavy Subfield		-0.002	-0.002	-0.002	-0.003	0.005	0.005
		(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)
Interactions with Network							
Chinese Investigator × Investigator with US Training			0.008*				0.008*
			(0.004)				(0.004)
Chinese Investigator × US First/Reprinted Cited Author			-0.028				-0.026
			(0.019)				(0.019)
Chinese Investigator × US Cited Author in Other Positions			0.003				0.002
			(0.010)				(0.010)
Chinese Investigator × Citation from Same Ethnicity			0.007†				0.008*
			(0.004)				(0.004)
Chinese Investigator × Investigator is an Editorial Author			-0.010				-0.009
			(0.018)				(0.018)
Interactions with Geographic Topical Clustering							
Chinese Investigator × Subfield Home Research Intensity				-0.065			-0.065
				(0.049)			(0.049)
Chinese Investigator × Subfield Foreign Research Intensity				0.023			0.023
				(0.017)			(0.017)
Chinese Investigator × Subfield USA Research Intensity				-0.024			-0.029
				(0.037)			(0.036)
Interactions with Investigator's Intellectual Focus							
Chinese Investigator × Importance of Subfield for Investigator					0.006*		0.006*
					(0.003)		(0.003)
Chinese Investigator × Importance of Investigator for the Subfield					-0.006		-0.005
					(0.034)		(0.035)
Chinese Investigator × Ellison/Glaeser Index of Scholarly Focus					0.083		0.067
					(0.105)		(0.104)
Interactions with Reputation							
Chinese Investigator × Retraction-heavy Subfield						-0.012†	-0.014*
						(0.007)	(0.007)
Mean of Dependent Variable	0.033	0.033	0.033	0.033	0.033	0.033	0.033
s.d. of Dependent Variable	0.177	0.177	0.177	0.177	0.177	0.177	0.177
China effect in s.d. units	-0.043	-0.038	-0.065	-0.040	-0.052	-0.035	-0.077
Adjusted R^2	0.078	0.083	0.083	0.083	0.083	0.083	0.083
No. of Investigators	531	531	531	531	531	531	531
No. of Cited Articles	15,122	15,122	15,122	15,122	15,122	15,122	15,122
No. of Citing Articles	68,963	68,963	68,963	68,963	68,963	68,963	68,963
No. of Citing/Cited Article Pairs	175,776	175,776	175,776	175,776	175,776	175,776	175,776

Notes: The dependent variable is an indicator variable that equals 1 if the related article cites the PI's article, and 0 otherwise. All regressions include fixed effects for rank bins of each citing article j with respect to its topic similarity to article i; fixed effects for the interaction of citing and cited article publication year; fixed effects for each CEM strata; fixed effects for the investigator's highest degree year and a investigator gender indicator variable (coefficients not reported). Standard errors in parentheses are two-way clustered at the investigator and strata level. $^{\dagger}p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table B.6: Summary statistics for control and treated articles matched on baseline covariates & a patent inventor status indicator

	Control articles (N= 8,810)					Treated articles (N= 7,279)					Difference	
	Mean	Median	Std. Dev.	Min	Max	Mean	Median	Std. Dev.	Min	Max	Mean	Median
Characteristics used for matching:												
Source Article Publication Year	2012.571	2013	3.768	2000	2018	2012.571	2013	3.768	2000	2018	0	0
Source Article No. of Authors	5.082	5	1.923	1	15	5.230	5	1.861	1	15	-0.148	0
Investigator PhD Degree Year	1993.737	1994	5.712	1958	2008	1993.940	1994	5.628	1952	2007	-0.203	0
No. of Debiased Article-to-Article Citations Outside US	15.701	10	18.723	0	328.856	15.251	9.867	17.345	0	345.264	0.450	0.133
Investigator is an Inventor	0.307	0	0.461	0	1	0.307	0	0.461	0	1	0	0
Characteristics not used for matching:												
No. of Total Article-to-Article Citations	20.554	14	23.133	0	387	28.400	18	32.565	0	580	-7.846**	-4*
No. of Article-to-Article Citations Outside US	17.711	11.750	20.058	0	333.300	25.969	16.607	29.793	0	521.303	-8.258**	-4.857*
No. of Debiased Patent-to-Article Citations Outside US	0.091	0	0.644	0	27.310	0.067	0	0.515	0	25	0.024*	0
US First/Reprinted Cited Author	0.019	0	0.135	0	1	0.006	0	0.076	0	1	0.013**	0
US Cited Author in Other Positions	0.035	0	0.183	0	1	0.013	0	0.115	0	1	0.022**	0
Retraction-heavy Subfield	0.027	0	0.163	0	1	0.040	0	0.196	0	1	-0.013**	0
Subfield Foreign Research Intensity	15.940	12.961	14.777	0	170.860	13.150	10.808	11.946	0	148.523	2.790**	2.153*
Subfield Home Research Intensity	1.507	0.710	2.288	0	38.926	10.752	7.704	11.512	0	158.569	-9.245**	-6.994*
Subfield USA Research Intensity	3.739	2.511	4.422	0	64.233	3.825	2.514	4.731	0	76.577	-0.086	-0.003
Importance of Investigator for the Subfield	0.063	0.038	0.078	0	1	0.053	0.031	0.069	0	1	0.010**	0.007*
Importance of Subfield for Investigator	0.111	0.045	0.285	0	10	0.113	0.036	0.328	0	7.333	-0.002	0.009*
Ellison/Glaeser Index of Scholarly Focus	0.016	0.014	0.014	-0.026	0.270	0.014	0.013	0.011	-0.020	0.142	0.002**	0.001*
Female Investigator	0.042	0	0.200	0	1	0.047	0	0.212	0	1	-0.005	0
Investigator with US Training	0.580	1	0.494	0	1	0.462	0	0.499	0	1	0.118**	1
Investigator from English-speaking Country	0.329	0	0.470	0	1	0	0	0	0	0	0.329**	0
Investigator is Editorial Author	0.036	0	0.187	0	1	0.010	0	0.101	0	1	0.026**	0
Investigator Publication Stock	86.758	74	63.579	0	1,068	99.953	82	75.845	0	482	-13.195**	-8*
Investigator Article-to-Article Citation Stock	1,676.827	1033	1,970.622	0	25,629	1,550.776	1004	1,700.989	0	12,811	126.051**	29*
Investigator Patent Stock	2.666	0	8.442	0	144	1.954	0	8.107	0	102	0.712**	0
Investigator Patent-to-Patent Citation Stock	15.531	0	78.958	0	1,581	4.481	0	17.065	0	185	11.050**	0

Notes: The stock of publications and citations for each investigator are assessed at the end of the year prior to the year of publication for each source article. The baseline set of matching covariates includes journal, publication year, number of authors (four bins: 1-3, 4-6, 7-9, 10 or more authors), home-debiased article-to-article citations outside US (six bins: $0-25^{th}$ percentile; $25-50^{th}$ percentile; $50-75^{th}$ percentile; $75-95^{th}$ percentile; $95-99^{th}$ percentile; and the top percentile), and the investigator PhD degree year (in three-year bins). In addition, the two groups of articles are matched on patent inventor status, i.e., an indicator variable equal to one if the PI has applied for at least one patent prior to the publication year of source article.

Table B.7: Summary statistics for control and treated articles matched on baseline covariates & investigator patent stock

	Control articles (N= 7,211)					Treated articles (N= 6,237)					Difference	
	Mean	Median	Std. Dev.	Min	Max	Mean	Median	Std. Dev.	Min	Max	Mean	Median
Characteristics used for matching:												
Source Article Publication Year	2,012.391	2013	3.802	2000	2018	2,012.391	2013	3.802	2000	2018	0	0
Source Article No. of Authors	5.009	5	1.844	1	15	5.158	5	1.781	1	15	-0.149	0
Investigator PhD Degree Year	1,993.796	1994	5.641	1958	2008	1,993.988	1994	5.571	1952	2007	-0.192	0
No. of Debiased Article-to-Article Citations Outside US	15.478	9.930	18.641	0	328.856	15.010	9.709	17.248	0	345.264	0.468	0.221
Investigator Patent Stock	1.738	0	7.441	0	144	1.675	0	7.891	0	90	0.063	0
Characteristics not used for matching:												
No. of Total Article-to-Article Citations	20.288	13	23.026	0	387	27.901	18	32.345	0	580	-7.613**	-5*
No. of Article-to-Article Citations Outside US	17.467	11.333	19.968	0	333.300	25.501	16.222	29.538	0	521.303	-8.034**	-4.889*
No. of Debiased Patent-to-Article Citations Outside US	0.089	0	0.658	0	27.310	0.065	0	0.500	0	25	0.024*	0
US First/Reprinted Cited Author	0.019	0	0.136	0	1	0.005	0	0.073	0	1	0.014**	0
US Cited Author in Other Positions	0.035	0	0.184	0	1	0.012	0	0.110	0	1	0.023**	0
Retraction-heavy Subfield	0.028	0	0.164	0	1	0.039	0	0.193	0	1	-0.011**	0
Subfield Foreign Research Intensity	15.944	12.897	14.782	0	170.860	13.183	10.778	12.109	0	148.523	2.761**	2.119*
Subfield Home Research Intensity	1.513	0.705	2.297	0	33.455	10.640	7.501	11.414	0	158.569	-9.127**	-6.796*
Subfield USA Research Intensity	3.771	2.511	4.449	0	64.233	3.805	2.503	4.735	0	76.577	-0.034	0.008
Importance of Investigator for the Subfield	0.063	0.039	0.078	0	1	0.055	0.032	0.071	0	1	0.008**	0.007*
Importance of Subfield for Investigator	0.119	0.049	0.304	0	10	0.121	0.038	0.351	0	7.333	-0.002	0.011*
Ellison/Glaeser Index of Scholarly Focus	0.017	0.014	0.015	-0.026	0.270	0.014	0.013	0.011	-0.020	0.142	0.003**	0.001*
Female Investigator	0.040	0	0.197	0	1	0.038	0	0.192	0	1	0.002	0
Investigator with US Training	0.581	1	0.493	0	1	0.443	0	0.497	0	1	0.138**	1
Investigator from English-speaking Country	0.325	0	0.469	0	1	0	0	0	0	0	0.325**	0
Investigator is Editorial Author	0.036	0	0.186	0	1	0.009	0	0.094	0	1	0.027**	0
Investigator is an Inventor	0.192	0	0.394	0	1	0.192	0	0.394	0	1	0	0
Investigator Publication Stock	83.350	71	62.728	0	1,068	97.746	80	75.292	0	482	-14.396**	-9*
Investigator Article-to-Article Citation Stock	1,581.116	925	1,966.909	0	25,629	1,469.358	928	1,636.790	0	12,811	111.758**	-3
Investigator Patent-to-patent Citation Stock	10.573	0	69.895	0	1,581	3.775	0	15.864	0	185	6.798**	0

Notes: The stock of publications and citations for each investigator are assessed at the end of the year prior to the year of publication for each source article. The baseline set of matching covariates includes journal, publication year, number of authors (four bins: 1-3, 4-6, 7-9, 10 or more authors), home-debiased article-to-article citations outside US (six bins: $0-25^{th}$ percentile; $25-50^{th}$ percentile; $50-75^{th}$ percentile; $75-95^{th}$ percentile; $95-99^{th}$ percentile; and the top percentile), and the investigator PhD degree year (in three-year bins). In addition, the two groups of articles are matched on the indicator variable for the investigator's patent stock up to the year of publication of the source article, divided into four coarse bins: zero patents, exactly one patent, exactly two patents, and three or more patents.

Table B.8: Summary statistics for control and treated articles matched on baseline covariates, patent inventor status, and non-US patent-to-article citations

	Control articles (N= 8,072)					Treated articles (N= 6,843)					Difference	
	Mean	Median	Std. Dev.	Min	Max	Mean	Median	Std. Dev.	Min	Max	Mean	Median
Characteristics used for matching:												
Source Article Publication Year	2,012.793	2013	3.677	2000	2018	2,012.793	2013	3.677	2000	2018	0	0
Source Article No. of Authors	5.093	5	1.925	1	15	5.246	5	1.857	1	15	-0.153	0
Investigator PhD Degree Year	1,993.807	1994	5.616	1958	2008	1,994.011	1994	5.518	1952	2007	-0.204	0
No. of Debiased Article-to-Article Citations Outside US	14.702	9.621	17.229	0	328.856	14.251	9.422	15.397	0	223.677	0.451	0.199
No. of Debiased Patent-to-Article Citations Outside US	0.004	0	0.086	0	4.473	0.003	0	0.082	0	4	0.001	0
Characteristics not used for matching:												
No. of Total Article-to-Article Citations	19.269	13	21.265	0	387	26.681	17	29.589	0	404	-7.412**	-4*
No. of Article-to-Article Citations Outside US	16.664	11	18.505	0	333.300	24.478	16	27.231	0	392.157	-7.814**	-5*
US First/Reprinted Cited Author	0.019	0	0.136	0	1	0.006	0	0.075	0	1	0.013**	0
US Cited Author in Other Positions	0.035	0	0.183	0	1	0.013	0	0.111	0	1	0.022**	0
Retraction-heavy Subfield	0.028	0	0.164	0	1	0.039	0	0.194	0	1	-0.011**	0
Subfield Foreign Research Intensity	15.825	12.890	14.684	0	170.860	12.991	10.673	11.851	0	148.523	2.834**	2.217*
Subfield Home Research Intensity	1.498	0.710	2.267	0	33.455	10.789	7.716	11.568	0	158.569	-9.291**	-7.006*
Subfield USA Research Intensity	3.640	2.455	4.275	0	64.233	3.735	2.450	4.664	0	76.577	-0.095	0.005
Importance of Investigator for the Subfield	0.063	0.039	0.078	0	1	0.054	0.032	0.070	0	1	0.009**	0.007*
Importance of Subfield for Investigator	0.109	0.045	0.284	0	10	0.110	0.034	0.319	0	7.333	-0.001	0.011*
Ellison/Glaeser Index of Scholarly Focus	0.016	0.014	0.014	-0.026	0.270	0.014	0.013	0.011	-0.020	0.142	0.002**	0.001*
Female Investigator	0.042	0	0.201	0	1	0.047	0	0.211	0	1	-0.005	0
Investigator with US Training	0.579	1	0.494	0	1	0.463	0	0.499	0	1	0.116**	1
Investigator from English-speaking Country	0.329	0	0.470	0	1	0	0	0	0	0	0.329**	0
Investigator is Editorial Author	0.039	0	0.193	0	1	0.011	0	0.102	0	1	0.028**	0
Investigator Publication Stock	88.644	76	64.117	0	1,068	102.533	84	76.418	0	482	-13.889**	-8*
Investigator Article-to-Article Citation Stock	1,729.294	1080	1,997.227	0	25,629	1,602.790	1,067	1,722.249	0	12,811	126.504**	13*
Investigator Patent Stock	2.599	0	8.179	0	144	1.879	0	7.724	0	102	0.720**	0
Investigator Patent-to-Patent Citation Stock	14.798	0	76.915	0	1,581	4.481	0	17.241	0	185	10.317**	0

Notes: The stock of publications and citations for each investigator are assessed at the end of the year prior to the year of publication for each source article. The baseline set of matching covariates includes journal, publication yearnumber of authors (four bins: 1-3, 4-6, 7-9, 10 or more authors), home-debiased article-to-article citations outside US (six bins: $0-25^{th}$ percentile; $25-50^{th}$ percentile; $50-75^{th}$ percentile; $75-95^{th}$ percentile; $95-99^{th}$ percentile; and the top percentile), and the investigator's PhD degree year (in three-year bins). In addition, the two groups of articles are matched on patent inventor status (as in Table B.6), as well as home-debiased patent-to-article citations outside the US, divided into four coarse bins: zero citation, 1 citation, 2 citations, and 3 or more citations.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1903	Jonathan Colmer Mary F Evans Jay Shimshack	Environmental citizen complaints
1902	Ester Faia Gianmarco I.P. Ottaviano Saverio Spinella	Robot adoption, worker-firm sorting and wage inequality: evidence from administrative panel data
1901	Federico Esposito Fadi Hassan	Import competition, trade credit and financial frictions in general equilibrium
1900	Fabrizio Leone	Multinationals, robots and the labor share
1899	Marcus Biermann Elsa Leromain	The indirect effect of the Russian-Ukraine war through international linkages: early evidence from the stock market
1898	Stephen J. Redding	The economics of cities: from theory to data
1897	Dan A. Black Jeffrey Grogger Tom Kirchmaier Koen Sanders	Criminal charges, risk assessment and violent recidivism in cases of domestic abuse
1896	Stephan Maurer Guido Schwerdt Simon Wiederhold	Do role models matter in large classes? New evidence on gender match effects in higher education
1895	Ihsaan Bassier	Collective bargaining and spillovers in local labor markets
1894	Paul C. Cheshire Christian A.L. Hilber Piero Montebruno Rosa Sanchis-Guarner	(In)convenient stores? What do policies pushing stores to town centres actually do?
1893	Kohei Takeda	The geography of structural transformation: Effects on inequality and mobility

1892	Martina Magli	The spillover effect of services offshoring on local labor markets
1891	Cheti Nicoletti Almudena Sevilla Valentina Tonei	Gender stereotypes in the family
1890	Rui Costa Swati Dhingra Stephen Machin	New dawn fades: Trade, labour and the Brexit exchange rate depreciation
1889	Amanda Dahlstrand	Defying distance? The provision of services in the digital age
1888	Jan David Bakker Nikhil Datta Richard Davies Josh De Lyon	Non-tariff barriers and consumer prices: evidence from Brexit
1887	Stefania Albanesi Claudia Olivetti Barbara Petrongolo	Families, labor markets and policy
1886	Dzhamilya Nigmatulina	Sanctions and misallocation. How sanctioned firms won and Russia lost
1885	Hugo de Almeida Vilares Hugo Reis	Who's got the power? Wage determination and its resilience in the Great Recession
1884	Stephan Heblich Stephen J. Redding Hans-Joachim Voth	Slavery and the British Industrial Revolution
1883	Boris Hirsch Elke J. Jahn Alan Manning Michael Oberfichtner	The wage elasticity of recruitment

The Centre for Economic Performance Publications Unit
Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk
Website: http://cep.lse.ac.uk Twitter: @CEP_LSE