EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Catalina Cozariuc, Luc Laeven,
Alexander Popov

The rise of China in academic research

No 3241

Abstract

Analyzing more than 300,000 articles across 40 top-tier journals between 2000 and 2022, this study demonstrates that China's 2006 National Medium- and Long-Term Plan for the Development of Science and Technology catalyzed a surge in publication volume and citations, propelling China past the United States as the world's leading producer of scientific research. Controlling for national income, population, and human capital, we find these gains are concentrated in fields explicitly targeted by the government's plan—physics, chemistry, biology, and medicine—while fields excluded from the plan, such as mathematics and economics, show significantly less growth. Our findings suggest that targeted state-led investment can effectively drive scientific progress, at least within a centrally planned economy.

JEL classification: F63; H52; I28; O38; P27

Keywords: Research and development; China; international competitiveness; government spending

Non-technical summary

China's emergence as a global scientific power is one of the most significant shifts in the modern research landscape. We examine how China achieved that rise and whether government-led investment in science and technology played a decisive role. Using data on more than 300,000 articles published in 40 leading academic journals across six disciplines between 2000 and 2022, we study how China's research output changed after the launch of the country's 2006 National Medium- and Long-Term Plan for the Development of Science and Technology (NMLP).

We find that, following the launch of the NMLP, China experienced a dramatic increase in both the quantity and quality of scientific research. Measured by publications in top journals, China's output rose sharply after 2006 and eventually surpassed that of the United States in several major scientific fields. The increase was especially large in disciplines that the Chinese government explicitly targeted in its national strategic plan, including physics, chemistry, biology, and medicine. On average, publications per capita by China-based researchers increased by about 17 percent relative to the rest of the world after the NMLP was introduced. Citation counts—a measure of research influence and quality—rose even more strongly. In fields directly targeted by the policy, publication output increased by roughly 26 percent relative to non-targeted fields, and the gap widened steadily over time. By 2022, publication output in targeted disciplines was roughly 85 percent higher than in the comparison group. By contrast, fields that were not major priorities under the plan—particularly economics and, to a lesser extent, mathematics—showed much slower growth.

Our analysis challenges the conventional assumption that centralised, state-led research systems are inherently inefficient or incapable of producing frontier innovation. Many observers had argued that China's top-down approach would encourage quantity over quality and lead to wasteful allocation of resources. Instead, the evidence suggests that sustained and strategically targeted public investment can substantially expand a country's scientific capacity and global influence.

The policy implications are considerable. First, our evidence points to government policy playing a major role in shaping scientific development, especially when investment is sustained, coordinated, and concentrated in priority areas. China's experience indicates that latecomer countries can rapidly build globally competitive research systems through deliberate public strategy. Second, our paper highlights the growing importance of scientific competition in the global economy. China's expanding research capacity poses a challenge to the long-standing dominance of the United States and Europe, implying that advanced economies may need to rethink their own science and innovation policies in order to remain competitive. Scientific leadership is clearly not fixed, and national investments in education, research infrastructure, and targeted innovation policy can significantly alter the global balance of knowledge production.

1 Introduction

For much of modern history, academic research has been dominated by European and American scientists (Adas 1989, 2008). What has received surprisingly little attention in the economics literature is that this has fundamentally changed in recent times with the rise of China as an economic powerhouse. Since China began to open its economy in 1978, it has experienced an average growth rate in GDP of close to 10 percent, and today it is the second largest economy in the world. Economic growth in China has been spurred by large-scale public investment in education and research, with the government spending immense resources on research and development (Van Noorden 2014; Sun and Cao 2014). Through a series of National Development Plans, China has consciously transformed itself into one of the world's leading research nations, dominating research output in many fields (Xie et al. 2014; Brainard and Normile 2022). Today, China counts over 3,000 universities and more than 1,200 highly cited researchers (Adams et al. 2023), having overtaken the United States in 2022 in terms of number of scientific publications (National Science Board of the National Science Foundation 2024).

A major impetus for the development of a domestic research capacity came in 2006 when the Chinese government launched its National Medium- and Long-Term Plan (NMLP) for the Development of Science and Technology (Sun and Cao 2021). This ambitious plan envisioned to transform China into an innovation-oriented economy by 2050, by directing scientific research in several technical fields of strategic importance and high-growth potential. The plan also included investments in mega industrial projects to reap synergies between research and development and practical experience, and reforms of the patent law to encourage the domestic patenting of ideas.

It is a priori not clear whether such government-led research and development initiatives are successful. On the one hand, technological progress is a key driver of long-term economic growth (Romer 1990; Aghion and Howitt 1992), and there is evidence that government spending can successfully spur innovation (Howell 2017). On the other hand, there is evidence of resource misallocation and rent-seeking in China, including in research and development expenditures (König et al. 2022; Hsieh and Klenow 2009; Wei et al. 2017; Young 2000) and many Western scholars hold the view that top-down centralized planning in innovation is bound to fail (e.g., Abrami et al. 2014).

In this paper, we examine how the rise of China has altered the publication of academic research across countries, across scientific fields, and over time. In which fields have Chinese scholars become dominant? What has been the role of government policies in bringing about this change? For identification, we exploit a shift in the supply of academic research caused by the launch of the 2006 NMLP in China. We focus our analysis on those fields that were targeted by the NMLP (physics, chemistry, biology, and medicine) and use mathematics and economics as the control group. We collect new bibliometric data on the global publication output in the top-10 academic journals in each of these academic fields from Elsevier's Scopus database, together with the citation count of each article, for the period 2000 to 2022. For each article, we also collect the affiliations and geographical locations of the authors. This allows us to assign a "nationality" to each paper, based on the location of the researcher.

Our findings point to some striking trends over time in how China has come to dominate many of these research fields. We find that for the period following the NMLP, China has seen an upward trend in its publications output and citations record across all fields. This upward trend for China is most pronounced in those research fields that were targeted by the NMLP program. In contrast, there is no similar trend in the field of economics, where the US and to a lesser extent Europe continue to dominate.

Our findings support the notion that research and development spending by the Chinese government has been highly successful in generating high-quality research in targeted research areas. This challenges the view held by many Western scholars prior to China's rise that innovation cannot be successfully led by centralized governments from the top down. Our analysis does not take a stand on whether resources were allocated efficiently and on whether this rise in research performance is sustainable. This new reality with China having become one of the global leaders in scientific publications shows that it is possible to overcome entry barriers, even as a latecomer, and that global leadership in research can be challenged.

Our paper contributes to several strands of literature. First, our paper relates to the literature on innovation and economic growth in China. This literature has focused on the growth of R&D and patenting of technologies in China's rise to technological power (Bergeaud and Verluise 2022) and the implications of such innovation for economic growth (Song et al. 2011). We contribute to this literature by analyzing the impact of government policy on China's academic performance in terms of publishing research.

Second, our paper relates to the literature on the production of academic research (see Stephan 1996 for a review). This literature has focused on such issues as the drivers of publication success (Aydin et al. 2023; Jones 2009; Azoulay et al. 2010) and biases in citations (Wilhite and Fong 2012; Fong and Wilhite 2017). Few papers have focused on the academic performance of Chinese economists. An exception is Xie and Freeman 2023 who focus on researchers of Chinese origin that are active in the US or have worked in the US but have returned to China and find that in both cases Chinese-born researchers outperform others in terms of research productivity. We contribute to this literature by analyzing how government spending can spur the production of academic research.

Third, our paper speaks to the literature on innovation and economic growth (e.g., Romer 1986, 1990). Within this literature, there is closely related work on the role of government policy in fostering innovation and research. For instance, Gross and Sampat 2023 show that large-scale government spending in the United States on research and development (R&D) through the Office of Scientific Research and Development (OSRD) shortly after the World War II had a major and lasting impact on US innovation, including through the formation of knowledge clusters. Our paper contributes to this literature by showing that public spending can spur the production of high-quality, innovative research in academia.

Our paper proceeds as follows. Section 2 provides detailed background on the initiatives of the Chinese government to spur academic research in China and offers a theoretical framework for our research. Section 3 presents the data. Section 4 presents the empirical methodology. Section 5 presents the empirical results. Section 6 discusses the implications of our findings, and Section 7 concludes.

2 Institutional background and framework

China has experimented with a top-down approach to innovation and science since it started reforming and opening its economy in 1978. Initially, its strategy focused on learning from the West and catching up. China started by allowing students and scholars to move abroad, to gain knowledge and foster research cooperation. In 1985, it passed a resolution to strengthen the role of applied research at the expense of fundamental research. The 1989 student protests culminated in the temporary break of ties with many Western institutions and a reorientation away from more liberal views. These political changes came at the expense of research in humanities and the social sciences and favored research in Science, Technology, Engineering, and Mathematics (STEM) disciplines. In the 1990s, government policy focused on promoting applied research and industrial development. These policies were complemented with programs to develop research-oriented universities and to attract more students. Over this period, the number of university students more than doubled Braun Střelcová et al. 2022. In the early 2000s, China continued its promotion of research with a top-down setting of research topics in key technologies and industrial fields. However, it was not until 2006 that China changed its ambition from catching up to moving ahead of others.

In January 2006, in a clear break with the past, China set out a strategy to turn the country into a global science superpower by 2050 (Braun Střelcová et al. 2022). This strategy was articulated in China's National Medium- and Long-Term Plan for the Development of Science and Technology (State Council of the People's Republic of China 2006). This long-term strategy was to be implemented through five-year economic plans. The plan encompassed: 11 key areas for national economic and social development, with 68 priority topics; 16 major special projects; 27 cutting-edge technologies and 18 basic scientific issues in 8 technical fields; and 4 major scientific research plans, spanning the fields of: biotechnology and life sciences; quantum electronics and quantum computing; nanotechnology and nanomaterial science; and reproductive biology, stem cell research, and genetics.

More generally, in the NMLP, China outlined three groups of objectives concerning scientific research:

- **Basic research in specific scientific fields.** This group includes cutting-edge scientific issues that can advance basic science in China. It encompasses areas such as quantitative analysis and system integration of life processes, the structure of condensed matter, the deep structure of matter and large-scale physical laws governing the universe, interdisciplinary mathematical applications, earth system processes and resources, environmental and disaster-related effects, chemical processes for creating and transforming new substances, brain and cognitive sciences, as well as innovations in experimental techniques, observational methods, and scientific technologies and equipment.

- **Basic research concerning major national strategic needs.** This group targets research topics with a strategic, long-term importance for China's socioeconomic development. It includes research on the biological foundations of health and disease, genetic advances in agriculture, sustainable agricultural practices, human impacts on the earth system, global and regional environmental changes, disaster forecasting and control, sustainable energy technologies, material design and manufacturing under extreme conditions, aerospace engineering, and the scientific basis for developing advanced information technologies.

- **Major scientific research plans.** These initiatives align with global scientific trends and aim to bolster China's international competitiveness. Key focus areas include protein research, quantum control research, nanotechnology, and reproductive sciences.

Thus, the fields broadly covered the major natural sciences (physics, chemistry, biology, and medicine). Less in focus were the related major sciences of mathematics and economics.

Under this ambitious growth plan, China's research and development expenses grew rapidly, almost doubling by 2010. Initially, the rapid growth was met with skepticism, with many observers questioning the effectiveness of the top-down approach, which they argued was creating a system that valued quantity over quality and was not conducive to innovation (Abrami et al. 2014). When Xi Jinping became China's president in 2013, he prioritized science, technology, and innovation policies as a cornerstone of the country's development strategy. Under his leadership, China established new funding organizations and research agencies to improve the efficiency with which research funds were allocated, and reduced its reliance on Western technology under the "Made in China 2025" initiative. China's international reputation in research started to rise, with the first Nobel prize in the natural sciences for a China-based Chinese scientist (Tu Youyou in 2015). The 13th Five Year Plan (2016-2020) set innovation-driven development as its central strategy, and by 2020, total research and development spending had reached the equivalent of about 378 billion US dollars per annum, or 2.4% of GDP. However, this strategy was perceived as coming at the expense of fundamental research. In response, the 14th Five Year Plan (2021-2025) raised the target for basic research to at least 8 percent of total research and development funding. By then, China had already become the top producer of academic research in many STEM fields.

More recently, the evaluation of academic journals in China underwent a fundamental change, moving from government-led assessment by the Chinese Academy of Sciences (CAS) toward a privately operated entity (Huang et al. 2021; Basu 2026). This shift altered how the quality of journals is measured, focusing less on Western metrics such as impact factors and more on domestic strategic priorities, thus channeling research funding toward strategic priorities and encouraging research in these areas.

We use China as a case study of a state-driven research and development strategy. It is a priori not clear whether such government-led research and development initiatives are successful. On the one hand, government funding of research can spur innovation by relaxing financing constraints, intensifying competition, and funding more researchers (e.g., Howell 2017; Arrow 1962; Jacob and Lefgren 2011). Government funding can also boost research through demand effects, with the government demanding research on new technologies by funding certain industries. On the other hand, there is evidence of resource misallocation and rent-seeking in China, including in research and development expenditures (e.g., König et al. 2022; Hsieh and Klenow 2009; Wei et al. 2017; Young 2000).

In our empirical setup, we measure the overall net effect of these supply and demand channels, not the separate channels. We measure the net effect of China's research and development strategy on its research output by using a difference-in-differences strategy that compares the evolution over time of the difference in research output between China-based and non-China-based researchers. For identification, we exploit a shift in the supply of academic research caused by the launch of the 2006 NMLP in China to distinguish between fields that were targeted by the NMLP (physics, chemistry, biology, medicine) and fields that were not. We measure research output using the number of publications in top academic journals and associated citations. The application of a difference-in-differences methodology allows estimating the causal effect of a government-led funding program (the NMLP) on China's scientific research over time.

3 Data

We focus our analysis on research published in top academic journals in STEM-related fields. Research published in these journals is an indication of research quality and global impact. We start by ranking all journals from SCImago (2026) across six scientific fields: physics, chemistry, biology, medicine, economics, and mathematics. The ranking is based on the combined ranking of two key metrics: the SCImago Journal Rank (SJR) and the H-index. SJR accounts for the number of citations received by a journal and the prestige of the journals where the citations come from, while the H-index is used as a measure of scientific productivity and impact of individual scientists. This approach allows us to incorporate both the citation impact and the overall influence of the journal. The combined rank of each journal is computed by first calculating the sum of the SJR and H-index rank by field (where lower values denote better rankings), and then ranking this sum from lowest to highest by field. We exclude journals that mainly publish reviews from the ranking (such as the Journal of Economic Literature or Nature Reviews journals). Based on this combined ranking, we select the top-10 ranked journals in each of the six fields, for a total of 61 journals. Four of these journals cover both the fields of chemistry and physics. The complete list of journals that we consider in our analysis can be found in Table A1.

The data associated with these 61 journals is manually retrieved from the Scopus (2026) user interface. We extract information for all publications classified as "Article" in Scopus (2026), totaling 784,579 unique articles.[1] The extracted data include the article title, identifiers (EID and DOI), citation count at the time of extraction, page count, author names and affiliations, and journal-specific information.

We use the Python library pybliometrics of Rose and Kitchin (2019) to systematically extract each author's primary affiliation and country via the Abstract Retrieval API, using either the DOI or EID as input. For each article, the API returns the full

---

[1] An article is defined in Scopus (2026) as original research or an opinion piece. This category excludes books, chapters, editorials, errata, letters, notes, reviews, etc.

list of authors along with their affiliation IDs. When an author has multiple affiliations, only the first listed affiliation is retained. In addition, we extract funding information for each article, including the name of the funding agency, grant identifiers, and the country of the funding body. Further details are provided in Appendix B.

For the analysis, we focus on long-standing journals that are well known in the profession by excluding journals that were established after 2000. This leaves us with 40 journals with information on 595,281 articles. The final list of journals included in the analysis is shown in Table 1. Because the historical coverage in Scopus (2026) is incomplete prior to 2000 we restrict the sample period for our analysis to the period 2000 to 2022 to ensure adequate coverage. The final sample includes 327,628 unique articles published between 2000 and 2022.

Table 1: Top Journals by Field

Field	**Journals**
Biology	Cell, Nature Genetics, Nature Biotechnology, Molecular Cell, Nature Cell Biology, Nucleic Acids Research, Genome Biology
Chemistry	J. Am. Chem. Society, Angewandte Chemie Int'l Edition, Appl. Catalysis B: Environmental
Physics	Advanced Materials, Materials Today, Physical Review Letters
Physics/Chemistry	Advanced Functional Materials
Medicine	New England J. Medicine, Nature Medicine, The Lancet, Immunity, Ca-A Cancer J. for Clinicians, Lancet Oncology, Nature Immunology, Annals of Oncology
Mathematics	Ann. Math., Ann. Stat., J. Am. Math. Soc., Invent. Math., Publ. Math. Inst. Hautes Etudes Sci., J. R. Stat. Soc. Ser. B, Commun. Pure Appl. Math., Biometrika, Acta Math.
Economics	Am. Econ. Rev., Quarterly J. Econ., J. Finance, J. Pol. Econ., Econometrica, Rev. Fin. Studies, J. Fin. Econ., Rev. Econ. Studies, Rev. Econ. Stat.

Notes: This table presents the journals included in the analysis, by field. Journals are ordered according to their combined ranking within field based on the SCImago rank and H-index. Sources: SCImago (2026) and Scopus (2026).

In addition, we collect country-specific variables to control for country traits that may influence the propensity to publish in academic journals. Population size and GDP per capita are sourced from World Bank (2023), while the education level (average years of schooling for the population aged 25 years and older) is obtained from UNESCO

(2023). We log-transform these variables for use in the regression models.

The citations data in Scopus (2026) are cumulative citations as of the time of data extraction. Specifically, cumulative citations are measured as of end-2025. For each article, we create a time series of citations by distributing the total number of citations evenly across all years since the publication of the article. Specifically, if $Citations_T$ is the total citations at the time of download T, and t_0 is the year of publication for the article, then the age-adjusted citation count of the article in every year $t \in \{2001, 2002, ..., 2022\}$ is $\frac{Citations_T}{(T-t_0)}$, where T=2025 and $T - t_0$ is the age of the publication in years. This linear transformation of the cumulative citations assumes that citations grow linearly with age.

We aggregate observations at the country-field-year level. Since many articles are co-authored by scholars from different countries, we weight each article according to the proportion of authors based in each country[2]. Only articles co-authored by up to 100 researchers are kept, representing 99% of all observations. We aggregate the observations for countries within the European Union, which is treated as a single entity. Finally, we balance the panel dataset to address cases where countries have no research output in certain fields and years. As a result, the final dataset is a balanced panel, with an equal number of observations per country and field across years. The final country-field-year dataset has a total of 22,632 observations. The descriptive statistics of the main variables are reported in Table A2.

In Tables 2 and 3, we report the fraction of articles produced by China-based and US-based researchers by field over distinct five year periods. We see a remarkable increase in the share of research output produced by China-based researchers over time, especially in the fields of chemistry and physics. The field of economics is an exception, with only a small increase in research originating from China and with US-based researchers maintaining a leading position in academic research output.

---

[2] For instance, if an article has four authors, with three based in the United States and one in China, the article is counted with weights of 0.75 for the United States and 0.25 for China.

Table 2: China Share of Articles by Field and Time Period (in %)

Field	2000-2005	2006-2010	2011-2016	2017-2022
Biology	0.782	2.486	4.973	11.905
Chemistry	2.411	6.767	15.066	36.353
Economics	0.218	0.554	1.043	1.780
Mathematics	1.153	2.398	3.784	5.330
Medicine	0.306	1.026	2.334	4.381
Physics	2.144	4.126	9.525	23.469

Table 3: US Share of Articles by Field and Time Period (in %)

Field	2000-2005	2006-2010	2011-2016	2017-2022
Biology	51.807	46.399	44.137	42.306
Chemistry	45.756	37.114	31.137	20.835
Economics	76.299	71.489	65.799	63.856
Mathematics	48.176	50.285	46.855	48.793
Medicine	41.614	46.500	45.395	47.007
Physics	34.974	32.137	28.791	24.234

4 Empirical strategy

Our empirical analysis consists of two parts. First, we study the evolution over time of the difference in research output between China-based and non-China-based academics. We consider both the quantity and quality of research output, measured in terms of publications and citations. Second, we extend this analysis by distinguishing between fields that were targeted by the NMLP (physics, chemistry, biology, medicine) and fields that were not. The application of a difference-in-differences methodology allows estimating the causal effect of the NMLP on China's scientific research over time.

To analyze the first effect, we employ the following difference-in-differences specification to explain the number of research publications:

$$
\begin{aligned}
ln\left(1+\frac{Publications}{Population}\right)_{i,j,t} &= \beta_1 China_i \times Post\_2006_t \\
&+ \beta_2 China_i + \beta_3 Post\_2006_t \\
&+ \gamma X_{i,t} + \Psi_{i,j} + \Phi_{j,t} + \varepsilon_{i,j,t}
\end{aligned}
\tag{1}
$$

where $Publications$ is the total number of academic publications in country i, field j, and year t, and $Population$ is the total population (in millions) in country i. We add a value of one to the dependent variable to avoid losing observations with zero publications when taking logs. $China_i$ is a dummy variable equal to one for publications by China-based researchers, and zero otherwise. $Post\_2006_t$ is a dummy variable equal to one for years starting from 2006 onward, and zero otherwise. The coefficient of interest, β_1, measures the overall effect of China, that is, the change in academic publications in China in the years after 2006, compared to the rest of the world before 2006.

Additionally, to analyze the evolving policy impact over time and observe the year when the China effect becomes significant, we estimate the following model with year-by-year coefficients:

$$
\begin{aligned}
ln\left(1+\frac{Publications}{Population}\right)_{i,j,t} &= \sum_{t=2001}^{2022} \beta_t China_i \times Year_t \\
&+ \gamma X_{i,t} + \Psi_{i,j} + \Phi_{j,t} + \varepsilon_{i,j,t}
\end{aligned}
\tag{2}
$$

where $Year_t$ is a dummy variable equal to one for year $t \in \{2001, 2002, ..., 2022\}$, and zero otherwise, with $t = 2000$ denoting the reference year. The coefficients of interest, β_t, for $t \in \{2001, 2002, ..., 2022\}$, measure the change in academic publications in the years 2001 to 2022, relative to 2000, for China compared to the rest of the world.

We saturate the empirical model in Equation 1 and Equation 2 with country-field and field-year fixed effects. $\Psi_{i,j}$ captures the impact on academic publications of factors that are fixed at the country-field level, such as time-invariant differences in the quality of secondary education and field-specific comparative advantages. $\Phi_{j,t}$ controls for any factors that vary across fields and over time but are common to all countries, such as the recent growth in research activity in the field of biology. We also include a vector of country-specific time-varying controls $X_{i,t}$ which includes determinants of scientific prowess such as GDP per capita, population size, and average years of schooling.

To analyze the second effect, we extend the regression model in Equation 1 as follows:

$$\begin{aligned} ln\left(1+\frac{Publications}{Population}\right)_{i,j,t} &= \eta_1 China_i \times MLP_j \times Post\_2006_t + \eta_2 China_i \times MLP_j \\ &+ \eta_3 China_i \times Post\_2006_t + \eta_4 MLP_j \times Post\_2006_t \\ &+ \eta_5 China_i + \eta_6 MLP_j + \eta_7 Post\_2006_t \\ &+ \Psi_{i,j} + \Phi_{i,t} + \Theta_{j,t} + \varepsilon_{i,j,t} \end{aligned} \tag{3}$$

where MLP_j is a dummy variable that is equal to one for the fields of physics, chemistry, biology, and medicine, and to zero for the fields of mathematics and economics. All other variables are as before. The differential NMLP effect of interest is captured by the η_1 coefficient on the three-way interaction term.

To analyze the evolving effect of the NMLP policy over time, we estimate the model below:

$$
\begin{aligned}
ln\left(1+\frac{Publications}{Population}\right)_{i,j,t} &= \sum_{t=2001}^{2022} \eta_t China_i \times MLP_j \times Year_t \\
&+ \Psi_{i,j} + \Phi_{i,t} + \Theta_{j,t} + \varepsilon_{i,j,t}
\end{aligned}
\tag{4}
$$

The coefficients of interest, η_t, for $t = 2001, ..., 2022$, measure the change in academic publications in years 2001 up to 2022, relative to 2000, for China compared to the rest of the world, and for a scientific field targeted by the NMLP, relative to other fields.

We saturate the empirical models in Equation 3 and Equation 4 with country-field, country-time, and field-time fixed effects, denoted $\Psi_{i,j}$, $\Phi_{i,t}$ and $\Theta_{j,t}$. $\Psi_{i,j}$ accounts for any differences in the propensity to publish in top academic journals that vary at the country-field level, such as Germany traditionally being strong in the field of chemistry. $\Phi_{i,t}$ controls for any time-varying factors at the country level, such as national wealth or economic growth. Finally, $\Theta_{j,t}$ controls for field-specific time-varying factors that are common to all countries, such as the growth in research activity over our sample period in the field of biology.

To analyze the effects on impact of research, as measured in terms of number of citations, we replace the dependent variable in the equations above with:

$$
ln\left(1+\frac{Citations}{Population}\right)_{i,j,t}
$$

where *Citations* is the age-adjusted citation count, summed over all articles in country i, field j, and year $t \in \{2001, 2002, ..., 2022\}$.

Specifically, we estimate the following specification to gauge the average effect of the NMLP on the relative performance of China-based researchers across all fields over time:

$$\begin{aligned}ln\left(1+\frac{Citations}{Population}\right)_{i,j,t} &= \beta_1 China_i \times Post\_2006_t \\ &+ \beta_2 China_i + \beta_3 Post\_2006_t \\ &+ \gamma X_{i,t} + \Psi_{i,j} + \Phi_{j,t} + \varepsilon_{i,j,t}\end{aligned} \tag{5}$$

and we estimate the following difference-in-difference-in-differences (DDD) specification to gauge the differential impact of the NMLP on the citation count of China-based researchers for fields that were targeted by the NMLP:

$$\begin{aligned}ln\left(1+\frac{Citations}{Population}\right)_{i,j,t} &= \eta_1 China_i \times MLP_j \times Post\_2006_t + \eta_2 China_i \times MLP_j \\ &+ \eta_3 China_i \times Post\_2006_t + \eta_4 MLP_j \times Post\_2006_t \\ &+ \eta_5 China_i + \eta_6 MLP_j + \eta_7 Post\_2006_t \\ &+ \Psi_{i,j} + \Phi_{i,t} + \Theta_{j,t} + \varepsilon_{i,j,t}\end{aligned} \tag{6}$$

The coefficients of interest are β_1 on the two-way interaction term in Equation 5 and η_1 on the triple interaction term in Equation 6.

5 Empirical results

5.1 All fields

We start by estimating Equation 1 to obtain the difference before and after 2006 in the research output between China and the rest of the world. Table 4 presents the difference-in-differences estimates in the number of publications and citations received between Chinese and non-Chinese research institutions, before and after the onset of the NMLP in 2006. The number of observations in column (2), where we include country-specific controls, is somewhat reduced because of missing data on schooling for a few countries.

The causal effect is economically and statistically significant for both the quantity of scientific output, as measured by the log of scaled publications, and the quality of scientific output, as measured by the log of scaled citations. Based on our preferred specification in column (2), a regression coefficient of 0.159 implies that the number of publications per million population is, on average, $e^{0.159} - 1 \approx 17.2\%$ higher for China-based publications after the onset of the policy in 2006, compared to the control group. Similarly, the citations received for these publications, adjusted for the age of the article and the population of the country, are, on average, $e^{0.876} - 1 \approx 140.1\%$ higher for the treatment group compared to the control group.

Results are robust to including no country-specific controls (column (1) and column (5)) and to limiting the sample to the main competing markets for China, namely the US (column (3) and column (7)) and the 27 member states of the European Union (column (4) and column (8)). The effect is larger compared to the US research institutions than compared to EU research institutions.

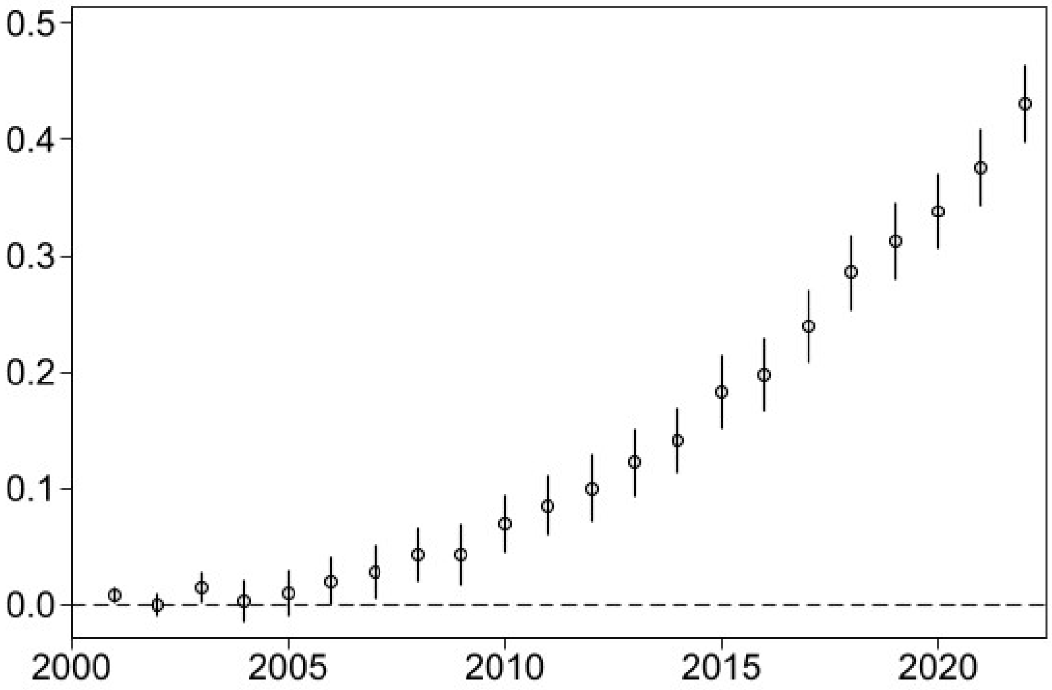
Next, we estimate Equation 2 to gauge the evolution over time of the differential effect. Figure 1 plots the regression coefficients, β_t, over time, together with confidence intervals. Standard errors are clustered at the country level. Before 2006, no significant differences were observed in the publication output between China and other countries. However, after 2006, a significant divergence occurs, with the publication gap steadily widening over time. The economic effect is substantial, with a coefficient of 0.431 in 2022, which implies that publications scaled by population are on average $e^{0.431} - 1 \approx 53.8\%$ higher for researchers based in China in 2022 than for the control group.

Table 4: Scientific Output in China and the Rest of the World: Difference-in-Differences Results

	Publications				Citations			
	(1) world	(2) world	(3) CN & US	(4) CN & EU27	(5) world	(6) world	(7) CN & US	(8) CN & EU27
China × Post_2006	0.159***	0.159***	0.096***	0.009***	0.884***	0.876***	0.524***	0.288***
	(0.008)	(0.011)	(0.000)	(0.000)	(0.022)	(0.038)	(0.000)	(0.000)
ln(GDP per capita)		-0.008				-0.038		
		(0.007)				(0.032)		
ln(Population)		0.029				0.060		
		(0.036)				(0.176)		
ln(Schooling)		-0.017*				-0.081**		
		(0.007)				(0.027)		
N	21666	18528	276	276	21666	18528	276	276
R-squared	0.898	0.904	0.969	0.963	0.869	0.878	0.968	0.957
Country × Field FE	yes	yes	yes	yes	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes	yes	yes	yes	yes

Notes: This table reports estimates from difference-in-differences (DiD) regressions evaluating the change in China's scientific output relative to the rest of the world after 2006. The dependent variable in columns (1)–(4) is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. The dependent variable in columns (5)–(8) is the natural logarithm of one plus population-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the interaction $China_i \times Post\_2006_t$, where $China_i$ is a dummy equal to one for China and $Post\_2006_t$ is a dummy for years $\geq$ 2006. Columns (2) and (6) include time-varying country-level controls: ln(GDP per capita), ln(Population), and ln(Schooling). Columns (3) and (7) restrict the sample to China and the US; Columns (4) and (8) restrict it to China and the EU27. All specifications include Country × Field and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Figure 1: Year-by-Year Estimates of China's Scientific Output Relative to the Rest of the World



Notes: This figure reports year-by-year estimates evaluating the change in China's scientific output relative to the rest of the world, as specified in equation (2). The dependent variable is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. Circles represent the regression coefficients β_t for the interaction $China_i \times Year_t$, with 95% confidence intervals shown by the vertical bars. The year 2000 serves as the reference period. The regression includes time-varying country-level controls: ln(GDP per capita), ln(Population), and ln(Schooling). The specification includes country-field and field-year fixed effects. Standard errors are clustered at the country level.

5.2 Distinguishing between fields

Next, we estimate Equation 3 to assess whether China's rise in publications is geared toward fields targeted by the NMLP (physics, chemistry, biology, and medicine), as opposed to fields that were not (mathematics and economics). Table 5 presents the difference-in-difference-in-differences estimates, while Figure 2 plots the year-by-year regression coefficients, along with confidence intervals. To make sure that the effect is tightly identified, we control for country-year, country-field, and field-year fixed effects.

The evidence suggests that there were no statistical differences between research output produced by Chinese and non-Chinese universities in NMLP compared to non-NMLP fields in the pre-treatment period, supporting the parallel trends assumption required for the difference-in-difference-in-differences specification to be a valid research design. After that, there is a clear and significant publication gap that widens over time, underscoring the sustained impact of the NMLP on China's scientific output.

Results are presented for both the number of publications (columns 1 to 3) and the number of citations (columns 4 to 6), and are qualitatively robust to focusing on the difference with the United States (columns 2 and 5) or Europe (columns 3 and 6). If anything, the results in terms of number of publications are stronger when contrasting China with the United States than when comparing to Europe, indicating that the rise of China in terms of number of publications has mostly come at the expense of the United States, which started out with a larger market than Europe. For citations, it is the other way around, with the effect being more pronounced for Europe than the United States.

The results in Table 5 show that the causal effect of the NMLP is both economically and statistically significant. Specifically, a regression coefficient of 0.234 in column (1) indicates that China's publication output in NMLP-targeted fields after 2006 increased by an average of $e^{0.234} - 1 \approx 26.4\%$ relative to the reference group.

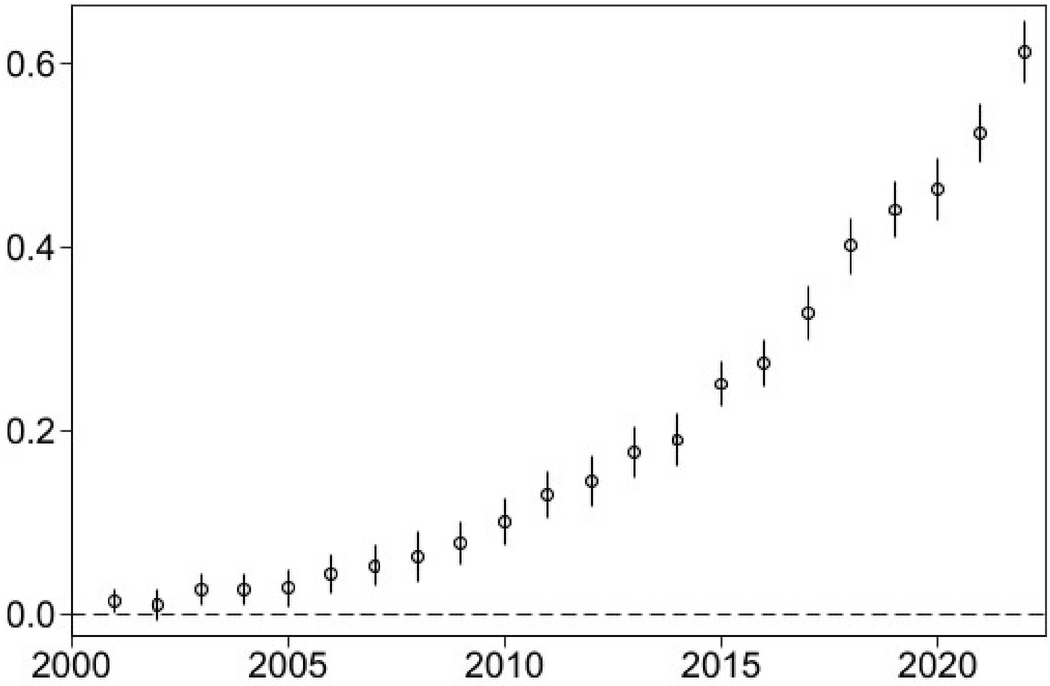
According to the year-by-year coefficients presented in Figure 2, the effect becomes more pronounced over time, with the regression coefficient of 0.613 in 2022 implying that the number of publications per million population is, on average, $e^{0.613} - 1 \approx 84.5\%$ higher for China-based researchers in targeted fields compared to the control group.

Table 5: Scientific Output in China by Field Priority: Triple-Difference Results for Targeted Fields

	Publications			Citations		
	(1) world	(2) CN & US	(3) CN & EU27	(4) world	(5) CN & US	(6) CN & EU27
China × MLP × Post_2006	0.234***	0.167***	0.074***	1.270***	0.796***	0.914***
	(0.010)	(0.000)	(0.000)	(0.025)	(0.000)	(0.000)
N	21666	276	276	21666	276	276
R-squared	0.928	0.981	0.979	0.912	0.984	0.977
Country × Field FE	yes	yes	yes	yes	yes	yes
Country × Year FE	yes	yes	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes	yes	yes

Notes: This table reports estimates from triple-difference (DDD) regressions evaluating the differential effect of the NMLP on targeted versus non-targeted fields in China. The dependent variable in columns (1)–(3) is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. The dependent variable in columns (4)–(6) is the natural logarithm of one plus population-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the triple interaction $China_i \times MLP_j \times Post\_2006_t$. MLP_j is a dummy equal to one if the field is a strategic objective of the NMLP (physics, chemistry, biology, or medicine) and zero otherwise (mathematics or economics). Columns (2) and (5) restrict the sample to China and the US; Columns (3) and (6) restrict it to China and the EU27. All specifications include Country × Field, Country × Year, and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Figure 2: Year-by-Year Estimates of China's Scientific Output in NMLP-Targeted versus Non-Targeted Fields



Notes: This figure reports year-by-year estimates evaluating the differential impact of the NMLP on targeted versus non-targeted fields in China, as specified in equation (4). The dependent variable is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. Circles represent the regression coefficients η_t for the triple interaction $China_i \times MLP_j \times Year_t$, with 95% confidence intervals shown by the vertical bars. The year 2000 serves as the reference period. The specification includes country-field, country-year, and field-year fixed effects. Standard errors are clustered at the country level.

5.3 Robustness tests

In this section, we report several robustness tests of our main empirical analysis. First, we control for whether the research received any funding support or grants. Academic research often receives support from public funding. China is no exception. To the extent that publicly funded research is more likely to get published (as shown by Jacob and Lefgren 2011) or there is country variation in the degree to which research receives public funding, not controlling for public funding may confound our analysis. We therefore extend our analysis by controlling for whether the article received any form of funding, whether the article received any form of funding from China, or whether the article received any form of funding from Chinese MLP agencies. Chinese MLP agencies are the four main government bodies in China that implement and fund MLP research priorities. They include the National Natural Science Foundation of China (established in 1986), the National Key Research and Development Program of China (established in 2016), the Ministry of Science and Technology of the People's Republic of China, and the National Basic Research Program of China (established in 2016). Among these four, the National Natural Science Foundation of China is the largest funder of academic research in China. Table 6 shows the distribution of articles in our sample that received any form of funding, China-based funding, or MLP funding. As shown in the table, 8.9% of all publications have MLP funding.

Table 6: Frequency of Articles with External Funding or Grants

Category	N	Percentage
Total articles	327,628	
Articles with any funding	189,946	58.0%
Articles with China funding	33,040	10.1%
Articles with MLP funding	29,322	8.9%
- National Science Foundation of China	28,506	8.7%
- National Key R&D Program of China	6,864	2.1%
- Ministry of Science and Technology of China	1,682	0.5%
- National Basic Research Program of China	1,548	0.5%

Tables A3 and A4 report results when we control in our main regression specifications for the percentage of articles at the country-field-year level that received *inter alia* any funding, China-based funding, or MLP-related funding. While we find that external funding tends to be positively associated with publication and citation output, our main results continue to hold.

Next, we consider whether results differ for papers where all the authors are from the same country. Tables A5 and A6 report results for our main regression specifications when restricting the sample to papers where all co-authors are affiliated to institutions in the same country. Results are qualitatively unaltered.

Another concern could be that we have scaled research output by the general population and not by the research-active population in a given field. The propensity to become a researcher may vary across countries and fields, and this may be correlated with the effect we are interested in. Moreover, results could be driven by changes in competition among scholars. For example, government funding could have resulted in more researchers, increasing competition among scholars (as in Jacob and Lefgren 2011). To proxy for the number of researchers in the country and in a given field, we use the number of researchers (and their country affiliation) with publications in a given field in our database. In Tables A7 and A8 we show that our main results continue to hold when we control for researcher density, the country-field specific number of researchers divided by total population (in millions) in the country. These results also indicate that research output at the field level is positively associated with the fraction of field-specific researchers in the population.

Similarly, results may be sensitive to scaling the dependent variable by the number of research institutions instead of the total population. In Tables A9 and A10 we present results when scaling publications and citations, respectively, by the total number of institutions. As proxy for the total number of institutions we use the number of unique affiliations in our dataset by country and year. Our main results continue to hold with this alternative form of scaling.

Another concern may be that there are potential positive knowledge transfers to researchers and institutions that were not originally targeted by government policy (as in Arrow 1962). Our results measure the total effects of the NMLP program, which could include spillover effects among NMLP fields and from NMLP to non-NMLP fields. To assess whether there were substantial spillover effects onto the control group, we re-run our main analysis separately for the subsamples of NMLP and non-NMLP fields. If the coefficient on $China \times Post\_2006$ is positive for the non-NMLP fields subsample, this would provide evidence in support of positive spillover effects. The results in Tables A11 and A12 do not provide any support for this. The results are insignificant for the non-NMLP subsample, and our main effect is entirely driven by the NMLP fields subsample.

Next, we evaluate the sensitivity of our results to potential bias induced by the $\ln(1 + y)$ transformation of the original data, with y being equal to (Publications/Population) or (Citations/Population), respectively. The tables A13 and A14 report estimates from Poisson Pseudo-Maximum Likelihood (PPML) estimation and inverse hyperbolic sine (IHS) regressions that are better capable of handling right-skewed dependent variables that contain zero values and in principle do not suffer from such potential transformation bias. For the PPML regressions, the dependent variable is the absolute number of publications (or citations). Standard errors are clustered at the country level. For the IHS regressions, we use the IHS transformation defined as $\text{asinh}(y) = \ln\left(y + \sqrt{y^2 + 1}\right)$, where y equals (Publications/Population) or (Citations/Population), respectively. The IHS transformation allows for zero and negative observations, while closely approximating $\ln(y)$ for large positive values. Our main results are robust to estimation using PPML estimation or the IHS specification, indicating that our main results are not biased due to the log transformation of the dependent variable.

6 Discussion

Our findings carry important implications for how governments design and finance research policy. First, they challenge the widespread skepticism that public investment—particularly in a centrally planned system—necessarily results in inefficiency or misallocation. The evidence indicates that strategically targeted government funding can meaningfully increase research output and international visibility, especially when resources are concentrated in clearly defined priority areas and supported through research centers of excellence.

Second, the results highlight the central role of the state in shaping the structure of a national research system aimed at converting fundamental research into commercially successful technologies. Well-documented US examples include the role of the Defense Advanced Research Projects Agency (DARPA) in nurturing the semiconductor industry and of the Small Business Administration (SBA) in developing a venture capital industry capable of financing the commercialization of science (Mazzucato 2015; Kuan and West 2023). In China's case, funding allocation is not a neutral administrative process but a powerful strategic instrument. Decisions about which disciplines and projects receive support have directed scientific capacity toward areas aligned with long-term industrial and technological objectives. This suggests that public funding can function as a lever of structural economic transformation, not merely as a supplement to private innovation.

Third, the field-specific nature of the effects underscores the importance of political and institutional context. STEM disciplines—particularly physics, chemistry, biology, and medicine—benefited directly from prioritization under national development strategies. These fields are closely tied to industrial upgrading and technological self-reliance, and their technical orientation may expose them to less political sensitivity. By contrast, economics and certain social sciences have not experienced comparable growth. Because these disciplines often involve critical analysis of policies and institutional arrangements, their development may be shaped by different institutional incentives and constraints. The divergence across fields illustrates how political economy considerations can influence the effectiveness of research policy. This puts in question whether in the foreseeable future, the field of economics and finance will play as central a role in China as it does in Western countries.

Fourth, the findings carry implications for global competition in science. China's expanding research capacity intensifies pressure on historically dominant scientific regions such as Europe and the United States. Sustaining leadership in the global knowledge economy will likely require adaptive policy responses, renewed investment, and institutional innovation. The evidence suggests that a state-driven research strategy—when coherent and sustained—can significantly shift the international balance of scientific output.

7 Conclusions

Using a dataset of more than 300,000 articles published in 40 leading scientific journals between 2000 and 2022, we examine the impact of China's 2006 National Medium- and Long-Term Plan for Science and Technology Development (NMLP). Prior to the reform, research output by China-based authors was broadly comparable to that of non-China-based researchers. Following the introduction of the NMLP, however, publications per capita increased significantly. On average, publications per million population were 17.2 percent higher for China-based researchers relative to the control group after 2006. The effect was concentrated in the scientific fields explicitly targeted by the reform, where output per capita rose by 26.4 percent relative to the control group. These findings remain robust when comparing China with the main competing scientific regions—the United States and Europe—and when controlling for income per capita, population size, and educational attainment.

The evidence therefore supports the view that government spending, when strategically directed, can promote measurable gains in scientific production and impact. At the same time, China's rise has not been uniform across disciplines. While STEM fields experienced strong expansion, economics has not seen a comparable surge in China-based research presence. This uneven development highlights both the targeted nature of the policy intervention and the continuing variation in China's comparative strengths across fields.

Taken together, the results document a substantial shift in the global research landscape and provide systematic evidence that coordinated state investment can play a decisive role in accelerating scientific development. At the same time, several open questions remain. It is unclear what distortions may arise from heavy reliance on public funding, how resources can be allocated most efficiently, and whether China can sustain its performance in the face of demographic aging and geopolitical tensions (Alfaro and Chor 2023). These uncertainties point to the need for continued evaluation of long-term policy outcomes.

Declarations

Conflict of interest. The authors declare no conflict of interest.

References

Abrami, R. M., Kirby, W. C., and McFarlan, F. W. (2014). Why China can't innovate. *Harvard Business Review*, 92(3):107–111.

Adams, J., Fry, R., Pendlebury, D., Potter, R., and Rogers, G. (2023). *Clarivate Global Research Report: China's research landscape.* Institute for Scientific Information, Clarivate.

Adas, M. (1989). *Machines as the Measure of Men: Science, Technology, and Ideologies of Western Dominance.* Cornell University Press.

Adas, M. (2008). *Western Dominance*, pages 2275–2279. Springer Netherlands.

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60(2):323–351.

Alfaro, L. and Chor, D. (2023). Global supply chains: The looming "great allocation". NBER Working Paper No. 31661, National Bureau of Economic Research.

Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In Nelson, R. R., editor, *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pages 609–626. Princeton University Press, Princeton, NJ.

Aydin, A., Yürük, S. E., Reisoğlu, İ., and Goktas, Y. (2023). Main barriers and possible enablers of academicians while publishing. *Scientometrics*, 128(1):623–650.

Azoulay, P., Graff Zivin, J. S., and Wang, J. (2010). Superstar Extinction. *Quarterly Journal of Economics*, 125(2):549–589.

Basu, M. (2026). Closure of China's influential journal ranking leaves academics reeling — what will take its place? *Nature News*. `https://www.nature.com/articles/d41586-026-01277-2`.

Bergeaud, A. and Verluise, C. (2022). The rise of China's technological power: The perspective from frontier technologies. CEP Discussion Papers 1876, London School of Economics and Political Science.

Brainard, J. and Normile, D. (2022). China rises to first place in one key metric of research impact. *Science*, 377(6608):799.

Braun Střelcová, A., Christmann-Budian, S., and Ahlers, A. L. (2022). The end of 'Learning from the West'? trends in China's contemporary science policy. Observations 6, Max Planck Institute for the History of Science.

Fong, E. A. and Wilhite, A. W. (2017). Authorship and citation manipulation in academic research. *PloS One*, 12(12):e0187394.

Gross, D. P. and Sampat, B. N. (2023). America, jump-started: World war ii r&d and the takeoff of the us innovation system. *American Economic Review*, 113(12):3323–56.

Howell, S. T. (2017). Financing innovation: Evidence from R&D grants. *American Economic Review*, 107(4):1136–1164.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4):1403–1448.

Huang, Y., Li, R., Zhang, L., and Sivertsen, G. (2021). A comprehensive analysis of the journal evaluation system in china. *Quantitative Science Studies*, 2(1):300–326.

Jacob, B. A. and Lefgren, L. (2011). The impact of research grant funding on scientific productivity. *Journal of Public Economics*, 95(9-10):1168–1177.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *Review of Economic Studies*, 76(1):283–317.

König, M., Storesletten, K., Song, Z., and Zilibotti, F. (2022). From imitation to innovation: Where is all that Chinese R&D going? *Econometrica*, 90(4):1615–1654.

Kuan, J. and West, J. (2023). Interfaces, modularity and ecosystem emergence: How darpa modularized the semiconductor ecosystem. *Research Policy*, 52(8).

Mazzucato, M. (2015). Building the entrepreneurial state: A new framework for envisioning and evaluating a mission-oriented public sector. Economics Working Paper Archive 824, Levy Economics Institute.

National Science Board of the National Science Foundation (2024). Publications output: U.S. trends and international comparisons. NSB-2023-33, Science and Engineering Indicators 2024.

Romer, P. M. (1986). Increasing Returns and Long-run Growth. *Journal of Political Economy*, 94(5):1002–1037.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5):S71–S102.

Rose, M. E. and Kitchin, J. R. (2019). Pybliometrics: Scriptable bibliometrics using a python interface to scopus. *SoftwareX*, 10:100263.

SCImago (2026). *Journal rankings.* SRG S.L. `https://www.scimagojr.com/journalrank.php/` (Accessed January 31, 2026).

Scopus (2026). *Database of peer-reviewed literature.* Elsevier, Amsterdam. `https://www.scopus.com/` (Accessed January 31, 2026).

Song, Z., Storesletten, K., and Zilibotti, F. (2011). Growing like China. *American Economic Review*, 101(1):196–233.

State Council of the People's Republic of China (2006). *National Medium- and Long-Term Plan for the Development of Science and Technology (2006–2020).* State Council of People's Republic of China. `https://www.gov.cn/gongbao/content/2006/content_240244.htm`.

Stephan, P. E. (1996). The economics of science. *Journal of Economic Literature*, 34(3):1199–1235.

Sun, Y. and Cao, C. (2014). Demystifying central government R&D spending in China. *Science*, 345(6200):1006–1008.

Sun, Y. and Cao, C. (2021). Planning for science: China's "grand experiment" and global implications. *Humanities and Social Sciences Communications*, 8(1):1–9.

UNESCO (2023). *UIS database.* UNESCO, Paris. https://data.uis.unesco.org/ (Accessed June 30, 2023).

Van Noorden, R. (2014). China tops Europe in R&D intensity. *Nature*, 505(7482):144–145.

Wei, S.-J., Xie, Z., and Zhang, X. (2017). From "Made in China" to "Innovated in China": Necessity, prospect, and challenges. *Journal of Economic Perspectives*, 31(1):49–70.

Wilhite, A. W. and Fong, E. A. (2012). Coercive citation in academic publishing. *Science*, 335(6068):542–543.

World Bank (2023). *World Development Indicators database.* World Bank, Washington, D.C. `https://databank.worldbank.org/source/world-development-indicators/` (Accessed June 30, 2023).

Xie, Q. and Freeman, R. B. (2023). Creating and connecting US and China science: Chinese diaspora and returnee researchers. NBER Working Paper No. 31306, National Bureau of Economic Research.

Xie, Y., Zhang, C., and Lai, Q. (2014). China's rise as a major contributor to science and technology. *Proceedings of the National Academy of Sciences*, 111(26):9437–9442.

Young, A. (2000). The razor's edge: Distortions and incremental reform in the People's Republic of China. *Quarterly Journal of Economics*, 115(4):1091–1135.

Supplemental Appendix

This Appendix contains supplementary material for the paper "The Rise of China in Academic Publications" by Catalina Cozariuc, Luc Laeven, and Alexander Popov.

Appendix A: Additional Tables and Robustness Tests

List of journals

Table A1: Journals Considered in the Analysis, by Field

Journal	**Field**	**Total Articles**		**Year**	**Rank**	**Included**
		total	since 2000			
Cell	Biology	15,858	7,401	1974	1	yes
Nucleic Acids Research	Biology	45,320	26,175	1974	8	yes
Nature Biotechnology	Biology	3,777	2,846	1989	3	yes
Nature Genetics	Biology	6,456	4,620	1992	2	yes
Molecular Cell	Biology	7,037	6,653	1997	5	yes
Nature Cell Biology	Biology	3,721	3,571	1999	5	yes
Genome Biology	Biology	3,626	3,626	2000	10	yes
Nature Methods	Biology	2,596	2,596	2004	4	no
Cell Metabolism	Biology	2,083	2,083	2005	7	no
Cell Host and Microbe	Biology	1,706	1,706	2007	9	no
Cell Stem Cell	Biology	1,543	1,543	2007	11	no
J. of the American Chemical Society	Chemistry	170,698	62,092	1879	2	yes
Angewandte Chemie - Int'l Edition	Chemistry	56,514	44,350	1962	4	yes
Applied Catalysis B: Environmental	Chemistry	14,023	13,446	1992	9	yes
Nature Chemistry	Chemistry	2,263	2,263	2009	5	no
ACS Energy Letters	Chemistry	2,574	2,574	2016	7	no
Chem	Chemistry	940	940	2016	11	no
Nature Catalysis	Chemistry	576	576	2018	9	no
Physical Review Letters	Physics	122,006	68,898	1958	11	yes
Advanced Materials	Physics	19,886	18,676	1989	2	yes
Materials Today	Physics	661	659	1999	8	yes
Nature Physics	Physics	2,959	2,959	2005	5	no
ACS Nano	Physics	18,068	18,068	2007	10	no
Nature Photonics	Physics	1,779	1,779	2007	4	no
Physical Review X	Physics	2,286	2,286	2011	8	no

Journal	Field	Total Articles		Year	Rank	Included
		total	since 2000			
Nature Electronics	Physics	417	417	2018	11	no
Advanced Functional Materials	Physics/Chemistry	17,400	17,400	2001	6	yes
Nature Materials	Physics/Chemistry	3,195	3,195	2002	1	no
Nature Nanotechnology	Physics/Chemistry	2,201	2,201	2006	3	no
Nature Communications	Physics/Chemistry	50,333	50,333	2010	6	no
The Lancet	Medicine	31,210	6,666	1826	3	yes
CA: A Cancer Journal for Clinicians	Medicine	1,429	226	1951	2	yes
New England Journal of Medicine	Medicine	18,955	10,135	1965	1	yes
Annals of Oncology	Medicine	7,689	6,418	1990	8	yes
Immunity	Medicine	3,910	3,176	1994	4	yes
Nature Medicine	Medicine	4,984	4,055	1995	2	yes
Nature Immunology	Medicine	3,062	3,062	2000	8	yes
The Lancet Oncology	Medicine	1,844	1,844	2000	6	yes
Cancer Cell	Medicine	1,969	1,969	2002	6	no
The Lancet Respiratory Medicine	Medicine	666	666	2013	10	no
Quarterly Journal of Economics	Economics	3,394	710	1888	2	yes
Review of Economic Studies	Economics	2,366	1,159	1933	7	yes
Journal of Finance	Economics	1,814	1,377	1946	3	yes
Journal of Political Economy	Economics	1,233	1,051	1969	4	yes
American Economic Review	Economics	3,098	2,154	1973	1	yes
Journal of Financial Economics	Economics	3,408	2,567	1974	7	yes
Econometrica	Economics	1,463	1,371	1978	5	yes
Review of Economics and Statistics	Economics	1,532	1,361	1980	9	yes
Review of Financial Studies	Economics	1,704	1,571	1996	6	yes
American Econ. J.: Applied Econ.	Economics	538	538	2009	10	no
Acta Mathematica	Mathematics	1,023	286	1887	9	yes
Biometrika	Mathematics	4,439	1,696	1908	8	yes
Communic. on Pure and Appl. Math.	Mathematics	2,213	1,071	1950	7	yes
Publ. Math. Inst. Hautes Études Sci.	Mathematics	342	145	1959	5	yes
Inventiones Mathematicae	Mathematics	4,373	1,586	1966	3	yes
J. of the American Math. Society	Mathematics	1,003	661	1988	3	yes
Annals of Mathematics	Mathematics	1,251	1,128	1996	1	yes
Annals of Statistics	Mathematics	2,674	2,287	1996	2	yes
J. of the Royal Stat. Society. Series B	Mathematics	1,180	1,031	1997	6	yes
Math. Programming Computation	Mathematics	232	232	2009	9	no
Annals of PDE	Mathematics	150	150	2015	9	no

Notes: This table presents the journals considered for analysis, including their corresponding fields, article counts, year of establishment, and combined ranking based on the SCImago Journal Rank (SJR) and H-index rank. The column "Included" indicates whether or not the journal is included in the final sample for analysis. Only journals established prior to 2001 are included in the analysis. Journals are ordered by field and by their year of establishment. Sources: SCImago (2026) and Scopus (2026).

Descriptive statistics

Table A2: Summary Statistics for the Country-Field Dataset, 2000-2022

	N	Min	Mean	Max	p25	p50	p75	SD
Publications	22,632	0.00	16.97	5146.92	0.00	0.00	0.08	135.90
Citations	22,632	0.00	2536.02	447817.22	0.00	0.00	6.58	20553.37
ln(Publications/Population)	21,666	0.00	0.11	3.42	0.00	0.00	0.00	0.37
ln(Citations/Population)	21,666	0.00	0.33	7.09	0.00	0.00	0.03	0.93
ln(Publications/Institutions)	21,990	0.00	0.05	1.39	0.00	0.00	0.00	0.13
ln(Citations/Institutions)	21,991	0.00	0.26	5.08	0.00	0.00	0.01	0.59
China	22,632	0.00	0.01	1.00	0.00	0.00	0.00	0.08
MLP	22,632	0.00	0.67	1.00	0.00	1.00	1.00	0.47
ln(GDP per capita)	21,300	4.70	8.28	12.39	7.03	8.21	9.33	1.58
ln(Population)	21,666	-3.45	1.95	7.26	0.86	2.16	3.35	2.10
ln(Schooling)	18,966	-1.47	1.83	2.63	1.57	2.06	2.36	0.71
Any funding (%)	22,632	0.00	15.90	100.00	0.00	0.00	0.00	33.23
China funding (%)	22,632	0.00	1.10	100.00	0.00	0.00	0.00	8.05
MLP funding (%)	22,632	0.00	0.83	100.00	0.00	0.00	0.00	6.96
Authors	22,632	0.00	47.17	10398.00	0.00	0.00	0.00	369.37
Institutions	22,632	0.00	72.75	5411.00	0.00	0.00	3.00	442.01
Researcher Density	21,666	0.00	0.74	71.85	0.00	0.00	0.00	3.90

Notes: This table reports descriptive statistics for all variables used in the analysis. *Publications* denotes the total number of articles published in country i, field j, and year t. *Citations* denotes the total number of citations (as of end-2025) received for articles published in country i, field j, and year t, adjusted for the age of the publication. *ln(Publications/Population)* and *ln(Citations/Population)* are computed as $\ln(1+x)$, where x represents the population-normalized publications and citations, respectively. *ln(Publications/Institutions)* and *ln(Citations/Institutions)* are similarly computed as $\ln(1+x)$ for publications and citations normalized by the number of research institutions. *China* is an indicator variable equal to one for China-based researchers and zero otherwise. *MLP* is an indicator variable equal to one if the field is a strategic objective of the NMLP (physics, chemistry, biology, or medicine) and zero for non-targeted fields (mathematics or economics). *ln(GDP per capita)*, *ln(Population)*, and *ln(Schooling)* are the natural logarithms of GDP per capita, total population, and average years of schooling, respectively. *Any funding (%)* represents the share of total articles in field j and year t with at least one reported funding source. *China funding (%)* and *MLP funding (%)* denote the share of total articles in field j and year t specifically funded by Chinese or NMLP-related funding agencies. *Authors* and *Institutions* represent the raw counts of researchers and research institutions, respectively. *Researcher Density* is defined as the total number of researchers per million population in country i, field j, and year t. The dataset covers 40 academic journals from 2000 to 2022. Sources: Scopus, World Bank, and UNESCO.

Controlling for research funding

Table A3: Difference-in-Differences Controlling for Research Funding

	Publications			Citations		
	(1)	(2)	(3)	(4)	(5)	(6)
China × Post_2006	0.150***	0.107***	0.096**	0.822***	0.673***	0.634***
	(0.012)	(0.026)	(0.029)	(0.036)	(0.078)	(0.088)
ln(GDP per capita)	-0.009	-0.010	-0.010	-0.042	-0.046	-0.047
	(0.007)	(0.007)	(0.007)	(0.027)	(0.029)	(0.030)
ln(Population)	0.029	0.027	0.026	0.057	0.051	0.047
	(0.032)	(0.033)	(0.033)	(0.137)	(0.166)	(0.167)
ln(Schooling)	-0.016**	-0.015**	-0.015*	-0.078***	-0.075**	-0.074**
	(0.006)	(0.006)	(0.006)	(0.022)	(0.025)	(0.025)
Any funding (%)	0.001***			0.004***		
	(0.000)			(0.000)		
China funding (%)		0.002*			0.009**	
		(0.001)			(0.003)	
MLP funding (%)			0.003*			0.011**
			(0.001)			(0.003)
N	18528	18528	18528	18528	18528	18528
R-squared	0.906	0.906	0.906	0.889	0.883	0.883
Country × Field FE	yes	yes	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes	yes	yes

Notes: This table reports estimates from difference-in-differences (DiD) regressions evaluating the change in China's scientific output relative to the rest of the world after 2006, controlling for various measures of research funding. The dependent variable in columns (1)–(3) is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. The dependent variable in columns (4)–(6) is the natural logarithm of one plus population-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the interaction $China_i \times Post\_2006_t$. All specifications include time-varying country-level controls: ln(GDP per capita), ln(Population), and ln(Schooling). Funding controls denote the share of total articles in field j and year t with any reported funding (*Any funding*), Chinese funding (*China funding*), or NMLP-specific funding (*MLP funding*). All specifications include Country × Field and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table A4: Triple-Difference Controlling for Research Funding

	Publications			Citations		
	(1)	(2)	(3)	(4)	(5)	(6)
China × MLP × Post_2006	0.240***	0.225***	0.220***	1.311***	1.228***	1.206***
	(0.009)	(0.012)	(0.013)	(0.024)	(0.034)	(0.042)
Any funding (%)	0.001***			0.003***		
	(0.000)			(0.000)		
China funding (%)		0.001			0.004*	
		(0.001)			(0.002)	
MLP funding (%)			0.001			0.005*
			(0.001)			(0.002)
N	21666	21666	21666	21666	21666	21666
R-squared	0.929	0.928	0.928	0.917	0.913	0.913
Country × Field FE	yes	yes	yes	yes	yes	yes
Country × Year FE	yes	yes	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes	yes	yes

Notes: This table reports estimates from triple-difference (DDD) regressions evaluating the differential effect of the NMLP on targeted versus non-targeted fields in China, controlling for various measures of research funding. The dependent variable in columns (1)–(3) is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. The dependent variable in columns (4)–(6) is the natural logarithm of one plus population-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the triple interaction $China_i \times MLP_j \times Post\_2006_t$. Funding controls denote the share of total articles in field j and year t with any reported funding (*Any funding*), Chinese funding (*China funding*), or NMLP-specific funding (*MLP funding*). All specifications include Country × Field, Country × Year, and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Single country papers

Table A5: Difference-in-Differences Excluding International Collaborations

	Publications		Citations	
	(1)	(2)	(3)	(4)
China × Post_2006	0.116***	0.119***	0.620***	0.626***
	(0.010)	(0.013)	(0.022)	(0.034)
ln(GDP per capita)		-0.007		-0.019
		(0.012)		(0.031)
ln(Population)		0.012		0.059
		(0.028)		(0.067)
ln(Schooling)		-0.008		-0.035
		(0.007)		(0.018)
N	12420	11406	12420	11406
R-squared	0.896	0.903	0.880	0.885
Country × Field FE	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes

Notes: This table reports estimates from difference-in-differences (DiD) regressions evaluating the change in China's scientific output relative to the rest of the world after 2006, using a sample restricted to single-country articles (i.e., excluding internationally co-authored publications). The dependent variable in column (1)–(2) is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. The dependent variable in columns (3)–(4) is the natural logarithm of one plus population-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the interaction $China_i \times Post\_2006_t$. Column (2) and (4) include time-varying country-level controls: ln(GDP per capita), ln(Population), and ln(Schooling). All specifications include Country × Field and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table A6: Triple-Difference Excluding International Collaborations

	Publications	Citations
	(1)	(2)
China × MLP × Post_2006	0.170***	0.918***
	(0.013)	(0.029)
N	12420	12420
R-squared	0.924	0.915
Country × Field FE	yes	yes
Country × Year FE	yes	yes
Field × Year FE	yes	yes
Clustered Std. Errors	yes	yes

Notes: This table reports estimates from triple-difference (DDD) regressions evaluating the differential effect of the NMLP on targeted versus non-targeted fields in China, using a sample restricted to single-country articles (i.e., excluding internationally co-authored publications). The dependent variable in column (1) is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. The dependent variable in column (2) is the natural logarithm of one plus population-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the triple interaction $China_i \times MLP_j \times Post\_2006_t$. All specifications include Country × Field, Country × Year, and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Controlling for researcher density

Table A7: Difference-in-Differences Controlling for Researcher Density

	Publications		Citations	
	(1)	(2)	(3)	(4)
China × Post_2006	0.159^{***}	0.138^{***}	0.884^{***}	0.835^{***}
	(0.008)	(0.006)	(0.022)	(0.029)
ln(GDP per capita)		0.001		-0.020
		(0.004)		(0.025)
ln(Population)		0.041		0.084
		(0.024)		(0.160)
ln(Schooling)		-0.009^{*}		-0.066^{**}
		(0.004)		(0.022)
Researcher Density		0.047^{***}		0.091^{***}
		(0.005)		(0.013)
N	21666	18528	21666	18528
R-squared	0.898	0.948	0.869	0.905
Country × Field FE	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes

Notes: This table reports estimates from difference-in-differences (DiD) regressions evaluating the change in China's scientific output relative to the rest of the world after 2006, controlling for the number of researchers per million population. The dependent variable in columns (1)–(2) is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. The dependent variable in columns (3)–(4) is the natural logarithm of one plus population-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the interaction $China_i \times Post\_2006_t$. Columns (2) and (4) include time-varying country-level controls: ln(GDP per capita), ln(Population), ln(Schooling), and *Researcher Density* (researchers per million population). All specifications include Country × Field and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table A8: Triple-Difference Controlling for Researcher Density

	Publications		Citations	
	(1)	(2)	(3)	(4)
China × MLP × Post_2006	0.234***	0.203***	1.270***	1.218***
	(0.010)	(0.009)	(0.025)	(0.028)
Researcher Density		0.046***		0.078***
		(0.006)		(0.015)
N	21666	21666	21666	21666
R-squared	0.928	0.957	0.912	0.925
Country × Field FE	yes	yes	yes	yes
Country × Year FE	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes

Notes: This table reports estimates from triple-difference (DDD) regressions evaluating the differential effect of the NMLP on targeted versus non-targeted fields in China, controlling for the number of researchers per million population. The dependent variable in columns (1)–(2) is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. The dependent variable in columns (3)–(4) is the natural logarithm of one plus population-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the triple interaction $China_i \times MLP_j \times Post\_2006_t$. Columns (2) and (4) include *Researcher Density* (researchers per million population) as an additional time-varying country-level control. All specifications include Country × Field, Country × Year, and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable scaled by number of institutions

Table A9: Difference-in-Differences with Institution-Normalized Output

	Publications		Citations	
	(1)	(2)	(3)	(4)
China × Post_2006	0.021***	0.030***	0.340***	0.326***
	(0.002)	(0.004)	(0.010)	(0.022)
ln(GDP per capita)		0.002		0.023
		(0.004)		(0.016)
ln(Population)		0.079***		0.235*
		(0.015)		(0.099)
ln(Schooling)		0.006		-0.001
		(0.004)		(0.023)
N	21990	17956	21991	17957
R-squared	0.659	0.672	0.698	0.706
Country × Field FE	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes

Notes: This table reports estimates from difference-in-differences (DiD) regressions evaluating the change in China's scientific output relative to the rest of the world after 2006, using an alternative normalization of output variables, based on the number of research institutions. The dependent variable in columns (1)–(2) is the natural logarithm of one plus institution-normalized publications in country i, field j, and year t. The dependent variable in columns (3)–(4) is the natural logarithm of one plus institution-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the interaction $China_i \times Post\_2006_t$. Columns (2) and (4) include time-varying country-level controls: ln(GDP per capita), ln(Population), and ln(Schooling). All specifications include Country × Field and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table A10: Triple-Difference with Institution-Normalized Output

	Publications (1)	Citations (2)
China × MLP × Post_2006	0.048***	0.531***
	(0.003)	(0.016)
N	21990	21991
R-squared	0.720	0.757
Country × Field FE	yes	yes
Country × Year FE	yes	yes
Field × Year FE	yes	yes
Clustered Std. Errors	yes	yes

Notes: This table reports estimates from triple-difference (DDD) regressions evaluating the differential effect of the NMLP on targeted versus non-targeted fields in China, using an alternative normalization of output variables, based on the number of research institutions. The dependent variable in column (1) is the natural logarithm of one plus institution-normalized publications in country i, field j, and year t. The dependent variable in column (2) is the natural logarithm of one plus institution-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the triple interaction $China_i \times MLP_j \times Post\_2006_t$. All specifications include Country × Field, Country × Year, and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Spillover effects

Table A11: Difference-in-Differences for NMLP Fields Only

	Publications		Citations	
	(1)	(2)	(3)	(4)
China × Post_2006	0.237***	0.237***	1.308***	1.297***
	(0.011)	(0.015)	(0.029)	(0.052)
ln(GDP per capita)		-0.011		-0.051
		(0.010)		(0.044)
ln(Population)		0.047		0.099
		(0.049)		(0.252)
ln(Schooling)		-0.026**		-0.122**
		(0.009)		(0.038)
N	14444	12352	14444	12352
R-squared	0.898	0.904	0.868	0.877
Country × Field FE	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes

Notes: This table reports estimates from difference-in-differences (DiD) regressions evaluating the change in China's scientific output relative to the rest of the world after 2006, restricting the sample to NMLP-targeted fields (physics, chemistry, biology, and medicine). The dependent variable in columns (1)–(2) is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. The dependent variable in columns (3)–(4) is the natural logarithm of one plus population-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the interaction $China_i \times Post\_2006_t$. Columns (2) and (4) include time-varying country-level controls: ln(GDP per capita), ln(Population), and ln(Schooling). All specifications include Country × Field and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table A12: Difference-in-Differences for Non-NMLP Fields Only

	Publications		Citations	
	(1)	(2)	(3)	(4)
China × Post_2006	0.003	0.002	0.038**	0.034*
	(0.004)	(0.004)	(0.014)	(0.015)
ln(GDP per capita)		-0.003		-0.013
		(0.004)		(0.014)
ln(Population)		-0.006		-0.017
		(0.014)		(0.050)
ln(Schooling)		0.002		0.001
		(0.004)		(0.011)
N	7222	6176	7222	6176
R-squared	0.864	0.876	0.856	0.860
Country × Field FE	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes

Notes: This table reports estimates from difference-in-differences (DiD) regressions evaluating the change in China's scientific output relative to the rest of the world after 2006, restricting the sample to non-NMLP fields (mathematics and economics). The dependent variable in columns (1)–(2) is the natural logarithm of one plus population-normalized publications in country i, field j, and year t. The dependent variable in columns (3)–(4) is the natural logarithm of one plus population-normalized, age-adjusted citations in country i, field j, and year t. The treatment is captured by the interaction $China_i \times Post\_2006_t$. Columns (2) and (4) include time-varying country-level controls: ln(GDP per capita), ln(Population), and ln(Schooling). All specifications include Country × Field and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Alternative specifications: PPML and IHS

Table A13: Difference-in-Differences with Alternative Specifications

	Publications		Citations	
	(1) Poisson	(2) IHS	(3) Poisson	(4) IHS
China × Post_2006	2.116***	0.198***	2.129***	1.105***
	(0.083)	(0.010)	(0.077)	(0.025)
N	13386	21666	13363	21666
R-squared		0.899		0.869
Pseudo R-squared	0.977		0.979	
Country × Field FE	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes

Notes: This table reports estimates from difference-in-differences (DiD) regressions evaluating the change in China's scientific output relative to the rest of the world after 2006, using alternative specifications to address potential bias from the $\ln(1+y)$ transformation of the dependent variable. Columns (1) and (3) report Poisson pseudo-maximum likelihood (PPML) estimates, where the dependent variable is the raw count of publications and citations, respectively. Columns (2) and (4) apply the inverse hyperbolic sine (IHS) transformation, defined as $\text{arcsinh}(y) = \ln\left(y + \sqrt{y^2+1}\right)$, where y equals Publications/Population and Citations/Population, respectively, in country i, field j, and year t. The treatment is captured by the interaction $China_i \times Post\_2006_t$. All specifications include Country × Field and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table A14: Triple-Difference with Alternative Specifications

	Publications		Citations	
	(1) Poisson	(2) IHS	(3) Poisson	(4) IHS
China × MLP × Post_2006	0.897***	0.294***	0.501***	1.604***
	(0.098)	(0.012)	(0.051)	(0.029)
N	9783	21666	9760	21666
R-squared		0.929		0.910
Pseudo R-squared	0.988		0.994	
Country × Field FE	yes	yes	yes	yes
Country × Year FE	yes	yes	yes	yes
Field × Year FE	yes	yes	yes	yes
Clustered Std. Errors	yes	yes	yes	yes

Notes: This table reports estimates from triple-difference (DDD) regressions evaluating the differential effect of the NMLP on targeted versus non-targeted fields in China, using alternative specifications to address potential bias from the $\ln(1+y)$ transformation of the dependent variable. Columns (1) and (3) report Poisson pseudo-maximum likelihood (PPML) estimates, where the dependent variable is the raw count of publications and citations, respectively. Columns (2) and (4) apply the inverse hyperbolic sine (IHS) transformation, defined as $\text{arcsinh}(y) = \ln\left(y + \sqrt{y^2+1}\right)$, where y equals Publications/Population and Citations/Population, respectively, in country i, field j, and year t. The treatment is captured by the triple interaction $China_i \times MLP_j \times Post\_2006_t$. All specifications include Country × Field, Country × Year, and Field × Year fixed effects. Standard errors (in parentheses) are clustered at the country level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Appendix B: Replication of Dataset

The dataset used in this study was originally retrieved manually through the Scopus user interface, consisting of 784,579 articles from 61 journals, identified at the article level. To enrich the dataset with author affiliations and funding information, the *Scopus Abstract Retrieval API* is used to systematically extract, for each article, the full list of authors and their affiliated countries, citation counts, and detailed funding data. To perform this extensive query, we relied on institutional access and enhanced API quotas provided by Scopus. The following instructions outline the dataset construction process.

API Configuration and Setup. To access and query the Scopus API, the Python library `pybliometrics` is installed and configured. After obtaining an Elsevier API key and an institutional token, the relevant modules from `pybliometrics` are imported.

```
from pybliometrics.scopus import config, AbstractRetrieval
```

Author and Affiliation Information Retrieval. The *Abstract Retrieval API* is used to obtain metadata associated with each publication, including the full list of authors, their affiliations, and the countries of their affiliated institutions. The query uses as input an article identifier, such as the **EID** (Scopus identifier) or **DOI**, and requests the full view of the abstract record. The main objective is to extract, for each article, the country of the primary affiliation for each co-author. The example below demonstrates a typical query:

```
eid = "insert_EID"
ab = AbstractRetrieval(eid, id_type="eid", view="FULL")
```

The `authors` attribute returns a list of namedtuples, one per author, where each entry includes an `affiliation` field containing the author's affiliation ID(s). When an author has multiple affiliations, only the first listed ID is retained. The `affiliation` attribute provides a mapping from affiliation IDs to institution names and countries, allowing us to resolve each author's primary affiliation country without a separate API call.

```
# Map affiliation IDs to countries
aff_lookup = {a.id: a.country for a in ab.affiliation}

# Extract the first affiliation country per author
countries = []
for author in ab.authors:
    first_id = int(author.affiliation.split(';')[0])
    countries.append(aff_lookup.get(first_id, ''))

author_count = len(ab.authors)
cited = ab.citedby_count
```

Funding Information Retrieval. In addition to author and affiliation data, the *Abstract Retrieval API* is used to extract funding information for each article. The `funding` attribute returns a list of namedtuples containing structured funding information, including the funding agency name, funding ID(s), and the country of the funding body.

```
funding = ab.funding or []
funding_agency = [f.agency for f in funding]
funding_id = [f.funding_id for f in funding]
funding_country = [f.country for f in funding]
```

Construction of article-country dataset. The process iterates through all articles in the dataset, systematically extracting and populating the relevant author-level, affiliation-level, and funding information. Each observation in the resulting dataset is identified at the article level and includes the associated countries of all co-authors as well as the funding details. If errors occur during data retrieval using the EID, the query is reattempted using an alternative article identifier (i.e., the DOI). If the issue persists, the data point is recorded as missing.

Aggregation at country-field level. The dataset described above provides a detailed overview of the geographic distribution of research contributions for each article. For the purpose of our analysis, the data is aggregated at the country-field level by assigning weights to each article based on the proportion of authors affiliated with institutions in each country. Each co-author contributes a weight of $1/N$ to their affiliated country, where N is the total number of authors. For instance, if an article has four authors, with three based in the United States and one in China, the article is counted with weights of 0.75 for the United States and 0.25 for China. This method allows for the construction of a comprehensive dataset that accurately reflects the country-level distribution of research contributions across different academic fields.

Acknowledgements

We would like to thank Jennie Bai, Charles Calomiris, Andrea Fabiani, Oded Galor, Seema Jayachandran, Hong Ru, Amit Seru, and seminar participants at the ECB for useful feedback and comments, and the ECB library team for help accessing the Scopus data. The opinions expressed herein are those of the authors and do not necessarily reflect those of the ECB or the Eurosystem.

Catalina Cozariuc
Imperial College London, London, United Kingdom; email: c.cozariuc25@imperial.ac.uk

Luc Laeven
European Central Bank, Frankfurt am Main, Germany; Centre for Economic Policy Research, London, United Kingdom;
email: luc.laeven@ecb.europa.eu

Alexander Popov
European Central Bank, Frankfurt am Main, Germany; CERGE-EI Foundation, Prague, Czech Republic; Centre for Economic Policy Research, London, United Kingdom; email: alexander.popov@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

PDF ISBN 978-92-899-7896-5 ISSN 1725-2806 doi:10.2866/1420467 QB-01-26-142-EN-N