https://doi.org/10.51408/issi2025_130

Country Self-Preference and National Research Systems: A Path to Independence or Isolation?

Jianjian Gao[1], Alexander J. Gates[2]

[1]psp2nq@virginia.edu, [2]agates@virginia.edu
School of Data Science, University of Virginia, Charlottesville, Virginia (USA)

Abstract

Scientific research is shaped by the interplay between national priorities and international collaboration, essential for advancing global knowledge. Although international collaboration improves research quality and impact, it may reduce national visibility and weaken domestic research networks. Conversely, self-reliant research ecosystems, often reflected in country self-citation rates, can promote research independence and address local challenges but may limit access to global resources. This study examines the balance between these dynamics using bibliometric data from OpenAlex, covering 264 countries from 1960 to 2023. We operationalize country self-preference by analyzing the proportion of citations countries give to their own work and measure international collaboration through the fraction of co-authored publications. The quality of national research ecosystems is assessed by the share of publications in top journals. Fixed-effect panel regression reveals that while international collaboration consistently boosts research quality, national self-preference also positively contributes when balanced effectively with collaboration. Our findings highlight the nuanced strategies nations employ to strengthen their research ecosystems, demonstrating that research independence and global collaboration can be complementary. This work provides actionable insights for policymakers seeking to optimize national scientific performance while fostering equitable and impactful international partnerships.

Introduction

Scientific research operates within a complex interplay of national priorities and international collaboration, both of which are critical for advancing global knowledge and innovation (Marginson, 2022). These dual imperatives often create tensions (Mormina, 2019; Harden-Davies and Snelgrove, 2020); while an interconnected global scientific ecosystem may heighten the overall productivity and efficiency of global science, it potentially diminishes national visibility and domestic research networks (Wagner et al., 2015). Furthermore, some countries, such as Iran, face exclusion from international collaboration due to geopolitical restrictions, forcing them to develop their research infrastructure and scientific capacity with limited external support and partnerships.

Here, we aim to uncover patterns in how nations navigate the trade-offs between fostering self-reliant research ecosystems and engaging in the global scientific enterprise. This question addresses a significant gap in current understanding, as existing literature predominantly focuses on promoting international collaboration without adequately considering its impact on national research ecosystems.

International collaboration has become a cornerstone of modern science, enabling the pooling of resources, expertise, and diverse perspectives to address complex global challenges (Wagner et al., 2001; Adams, 2012). Research consistently underscores

the advantages of cross-border scientific partnerships, as internationally co-authored publications tend to achieve higher citation rates and broader cross-disciplinary impact compared to domestic-only collaborations (Wagner and Jonkers, 2017; Adams, 2013; Glanzel and Schubert¨ , 2001). Over the past three decades, the scale and scope of international collaboration have expanded remarkably (Wagner and Leydesdorff, 2005; Leydesdorff and Wagner, 2008; Chen et al., 2019). Recognizing these benefits, policymakers increasingly prioritize international partnerships in research funding strategies (Katz and Martin, 1997). As Wagner et al. (2015) describe, the global research network is emerging as a new organizational structure that complements—and in some cases supersedes—traditional national systems. For developing countries, international collaboration often serves as a critical mechanism for building national scientific capacity (Harris, 2004). However, despite its many advantages, global collaboration networks remain unequal. Researchers from higher-income countries frequently dominate partnerships, shaping research agendas and benefitting disproportionately (Glanzel and Schubert¨, 2001). Furthermore, geopolitical tensions, funding limitations, and language barriers present significant obstacles to equitable participation in international science, underscoring the need for policies that foster more inclusive and sustainable collaboration frameworks.

Country self-citation offers valuable insights into national research ecosystems, reflecting the extent to which nations rely on and build upon their domestic scholarly contributions (Bakare and Lewison, 2017; Shehatta and Al-Rubaish, 2019; Baccini et al., 2019; Baccini and Petrovich, 2023). While often criticized as a sign of insularity or bias—potentially inflating metrics like the h-index and journal rankings—self-citation can also signify research independence and the ability to address local challenges, particularly in maturing scientific systems (Lariviere et al., 2018; Ladle et al., 2012). Various metrics, such as the self-citation rate and over-citation ratio, attempt to quantify this phenomenon, though recent approaches like fractional citation counts aim to reduce biases related to country size (Qiu et al., 2024). While self-citation often increases alongside international collaboration within countries, its prevalence varies globally, with higher rates in developing nations reflecting localized research priorities or limited visibility, whereas lower rates in developed countries signify greater integration into global networks (Baccini et al., 2019).

In this study, we critically examine these two pivotal dimensions that significantly influence the scientific performance of nations: country self-preference in citations and international collaboration. Specifically, we use bibliometric data from OpenAlex and operationalize country self-preference by analyzing the distribution of citations a country gives to itself relative to all other countries, employing the Area Under the Receiver Operating Characteristic Curve (AUC) and stratified bootstrap to control for a publication's journal. We next measure international collaboration through the fraction of publications involving international co-authors. Lastly, the quality of a nation's scientific ecosystem is measured by the proportion of its articles published in top journals. We then examine the intricate interplay between country self-preference and international collaboration in driving publications in top journals,

using fixed-effect panel regression to uncover their combined impact. Additionally, we explore pathways for strengthening scientific capacity, identifying level sets which reflect the trade-offs between international collaboration or bolstered domestic research infrastructure. Understanding these dynamics is critical for shaping global science policy, as it highlights the nuanced and emergent strategies adopted nations to optimize their research ecosystems and enhance their contributions to the global scientific enterprise. By identifying key drivers and trade-offs, this work provides actionable insights for policymakers to foster equitable and impactful global collaboration while supporting the sustainable development of national research systems.

Data and Methods

We leverage bibliometric data drawn from the OpenAlex bibliometric database in July 2022. We used all indexed "journal-article" and "proceedings-article" records listed as published after 1900 and excluded any publication that did not list an institutional address. Publications are associated with countries using the institutional addresses listed by the authors. We assign a full unit credit of a publication to every country of affiliation on the paper's author byline ("full counting"). In addition, we control for the influence of author self-citation (Aksnes, 2003) and institution self-citation (Wuestman et al., 2019) by removing all citations between publications that share at least one author or at least one affiliation.

We use data on national GDP and percentage of R&D investment from the World Bank to approximate the economic wealth and size of each country. The dataset covers 264 countries from 1960 to 2023.

Fraction with International Authorship

For each country in each year, we count the fraction of publications that share at least one authorship with at least one other country (international collaboration).

National citation preference

We fix a year y, a source country (citing country) s and a target country t (cited country). We then find all publications $n_{s,y}$ worldwide published in year y that also received citations from the source country's 5-year publications. Then, we focus on a target country t, identifying a subset of $n_{s,t,y}$ publications within $n_{s,y}$ publications. We produce 100 stratified bootstrap samples from the worldwide distribution such that the number of publications in each journal exactly matches the counts observed in the target country, thus controlling for both disciplinary differences in citation and one sense of scientific quality.

Finally, we use the Area Under the receiver-operator Curve (AUC) as a measure of the extent to which the cited country's publications ($n_{s,t,y}$) are randomly distributed throughout the citing country's ($n_{s,y}$) ranking. The AUC is a measure of the probability (between 0 and 1) that a randomly chosen publication from the cited country is ranked higher than a randomly chosen publication from any other country; a value of 1 reflects the cited country's publications are over-expressed towards the

top of the ranking, 0 occurs when the cited country's publications are under-expressed towards the bottom of the ranking, and 0.5 denotes a random distribution throughout the ranking. In this study, we set the source country *s* and the target country *t* to be the same country, and obtain the AUC for country self-preference (See Gates et al. (2024) for more details).

We can further quantify the statistical significance of the over/under-representation of a specific country in the citation counts due to the equivalence of the AUC and Mann-Whitney U statistic (DeLong et al., 1988; Sun and Xu, 2014).

Normalized Fraction of Top Journal Articles

To capture the quality of a nation's scientific ecosystem, we use a normalized measure for the fraction of their articles appearing in top journals. Specifically, for each journal in each year, we take the mean log of citations over 5 years to each of its articles. We then rank the journals with publications in each of the 252 subfields in OpenAlex, and take the top 50 journals by subfield; the union set over subfields represents our top journal selection. We then create a normalized measure by dividing by the total number of articles in those journals in that year, thus controlling for the variation in publication volume.

International collaboration and domestic research capacity

We first quantify the relationship between the scientific strength of a nation and the strength of its international collaboration or self-citation preferences. These phenomena, though often studied separately, are intricately linked to the broader dynamics of the scientific capacity of nations.

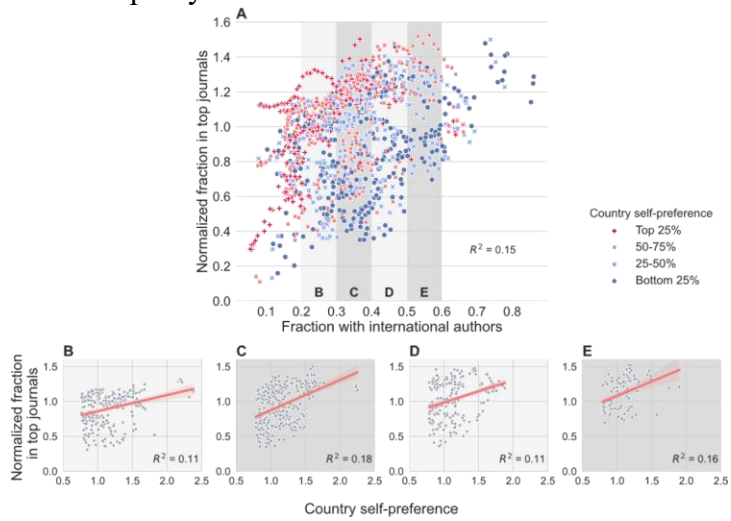

Figure 1: Panel A shows the relationship between the fraction of publications. Panels B-F detail conditional relationships within specific international publication fractions, with corresponding regression lines annotated by R-square (R^2). The shaded area around the regression line represents the 95% confidence interval.

As shown in Fig.1A, there is a strong positive correlation between the fraction of publications with international authors and the normalized fraction of publications in top journals, emphasizing the critical role of international collaboration in enhancing research impact. This corroborates the conclusion by Wagner and Jonkers (2017) that "open countries have strong science". However, the level of openness alone cannot fully explain the variation in high-impact publication rates. Among countries with similar levels of international collaboration, the normalized fraction of top articles varies significantly, suggesting that other factors contribute to research success. The impact of country self-preference, as a proxy of country's scientific independence, on top journal performance is not uniform. When examining countries with similar levels of internationalization (30-40% or 40-50% international authorship), a clearer positive linear trend emerges (Fig. 1B-E), suggesting that the relationship between self-preference and high-impact publications is more readily observable when comparing countries with similar degrees of international collaboration.

To further elucidate this relations, we use a two-way fixed effect panel regression model in which country self-preference and/or international authorship is used to predict the quality of national scientific outputs in the presence of several common covariates. The regression consistently shows a strong, positive, and highly significant effect across all models for international collaboration ($\beta = 0.1661 \sim 0.1824$, p-value < 0.01), underscoring its critical role in enhancing research quality. Country self-preference also shows positive and significant when controlling for other factors ($\gamma = 0.1021$, p-value < 0.05), suggesting that national autonomy in research can complement collaboration when balanced effectively. High self-preference might reflect a country's capacity for independent research, but its translation into impactful publications improves significantly when coupled with robust international partnerships.

Level-sets of scientific capacity

The interplay between international collaboration and research independence represents a fundamental tension in developing national research systems. Some developing countries rely heavily on international collaborations for training and knowledge-sharing (Harris, 2004), while others struggle with national infrastructure building to solve region-specific problems. This section aims to reveal the tension between these two strategies and their impact on the effectiveness of national research systems.

Fig.2 depicts the relationship between country self-preference and international collaboration as resources for producing high-quality research. For varying levels of scientific quality, indicated by the lines (Q1–Q4, from bottom 25% to top 25% according to the ranking of the normalized fraction of publications in top journals), can be interpreted as different patterns of resource utilization, where higher lines (e.g., Q4) represent more efficient or strategic combinations of these resources, yielding greater normalized fractions of publications in top journals.

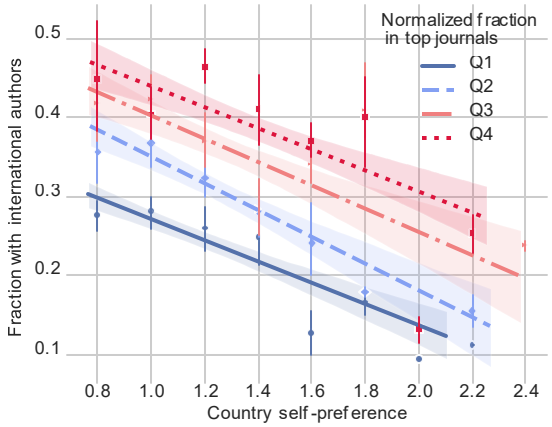

Figure 2: Balance between country self-preference (log of the z-score of the AUC) and international authorship. Shaded areas indicate 95% confidence intervals.

For a fixed quantile of scientific quality, there's a negative trend between country self-preference and the fraction of publications with international authors. This indicates that countries relying on domestic research capacity tend to have fewer external partnerships. This trend reveals the tension between the inclination towards building domestic infrastructure and promotion of international collaboration for national scientific capacity building.

More importantly, this figure demonstrates paths through which countries transition from a lower quality level to a higher one. Countries at lower levels (e.g., Q1) may over-rely on one resource without optimizing the balance due to economic constrains or geopolitical tensions, whereas those on higher levels demonstrate more efficient or strategic resource allocation. To ascend to higher levels, countries must enhance the weaker resources — whether by increasing international collaboration or strengthening domestic research capacity. This analysis underscores the importance of strategic resource utilization and equitable access to collaboration opportunities for achieving research impact. Fostering international partnerships should be complemented by policies to strengthen domestic research infrastructure, support independent researchers, and promote local innovation ecosystems.

Discussion

Our study reveals the complex interplay between international collaboration and country self-preference in scientific research, offering critical insights into national research ecosystem dynamics. Key observations highlight that both international cooperation and self-referential publication practices contribute positively to high-quality research output. The findings have profound implications for science policy. Nations can pursue diverse scientific capacity-building strategies: some may prioritize extensive international networks, while others may focus on strengthening domestic research infrastructures. The observed trends and trade-offs are relevant irrespective of their underlying cause—whether they arise as an emergent properties of the national scientific ecosystem or result from deliberate strategic policy

decisions by governments—highlighting their importance for understanding global research infrastructure. The limitations of this study include potential biases in the OpenAlex dataset and the complexity of measuring the quality of research through the fraction of publications in top journals. Future research could explore individual countries' trajectories of science capacity-building and policies driving the current landscapes in countries.

References

Adams, J. (2012). The rise of research networks. *Nature 490*(7420), 335–336.

Adams, J. (2013). The fourth age of research. *Nature 497*(7451), 557–560.

Aksnes, D. W. (2003). A macro study of self-citation. *Scientometrics 56*(2), 235–246.

Baccini, A., G. De Nicolao, and E. Petrovich (2019). Citation gaming induced by bibliometric evaluation: A country-level comparative analysis. *PLoS One 14*(9), e0221212.

Baccini, A. and E. Petrovich (2023). A global exploratory comparison of country self-citations 1996-2019. *Plos one 18*(12), e0294669.

Bakare, V. and G. Lewison (2017). Country over-citation ratios. *Scientometrics 113*(2), 1199–1207.

Chen, K., Y. Zhang, and X. Fu (2019). International research collaboration: An emerging domain of innovation studies? *Research Policy 48*(1), 149–168.

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson (1988, Sep). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics 44*(3), 837.

Gates, A. J., I. Mane, and J. Gao (2024). The increasing fragmentation of global science limits the diffusion of ideas. Available at arXiv: https://doi.org/10.48550/arXiv.2404.05861

Glanzel, W. and A. Schubert (2001). Double effort= double impact? A critical view at international co-authorship in chemistry." *Scientometrics 50*(2), 199–214.

Harden-Davies, H. and P. Snelgrove (2020). Science collaboration for capacity building: Advancing technology transfer through a treaty for biodiversity beyond national jurisdiction. *Frontiers in Marine Science 7*, 40.

Harris, E. (2004). Building scientific capacity in developing countries. *EMBO reports 5*(1), 7–11.

Katz, J. S. and B. R. Martin (1997). What is research collaboration? *Research policy 26*(1), 1–18.

Ladle, R. J., P. A. Todd, and A. C. Malhado (2012). Assessing insularity in global science. *Scientometrics 93*(3), 745–750.

Lariviere, V., K. Gong, and C. R. Sugimoto (2018). Citations strength begins at home. *Nature 564*(7735), S70– S70.

Leydesdorff, L. and C. S. Wagner (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics 2*(4), 317–325.

Marginson, S. (2022). "All things are in flux": China in global science. *Higher Education 83*(4), 881–910.

Mormina, M. (2019). Science, technology and innovation as social goods for development: Rethinking research capacity building from Sen's capabilities approach. *Science and Engineering Ethics 25*(3), 671–692.

Qiu, S., C. Steinwender, and P. Azoulay (2024, May). Paper tiger? Chinese science and home bias in citations. Working Paper w32468, National Bureau of Economic Research. Available at SSRN: https://ssrn.com/abstract=4833948.

Shehatta, I. and A. M. Al-Rubaish (2019). Impact of country self-citations on bibliometric indicators and ranking of most productive countries. *Scientometrics 120*(2), 775–791.

Sun, X. and W. Xu (2014, Nov). Fast implementation of Delong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Processing Letters 21*(11), 1389–1393.

Wagner, C. S., I. Brahmakulam, B. Jackson, A. Wong, and T. Yoda (2001). Science and technology collaboration: Building capacity in developing countries. *Document Number MR-1357.0-WB, RAND Corporation, Santa Monica, CA*.

Wagner, C. S. and K. Jonkers (2017, October). Open countries have strong science. *Nature 550*(7674), 32–33.

Wagner, C. S. and L. Leydesdorff (2005). Mapping the network of global science: Comparing international co-authorships from 1990 to 2000. *International journal of Technology and Globalisation 1*(2), 185–208.

Wagner, C. S., H. W. Park, and L. Leydesdorff (2015, July). The continuing growth of global cooperation networks in research: A conundrum for national governments. *PLoS One 10*(7), e0131816.

Wuestman, M. L., J. Hoekman, and K. Frenken (2019). The geography of scientific citations. *Research Policy 48*(7), 1771–1780.