MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany · Tel +49 (0) 3 81 20 81 - 0 · Fax +49 (0) 3 81 20 81 - 202 · www.demogr.mpg.de

MPIDR Working Paper WP 2023-038 I September 2023
https://doi.org/10.4054/MPIDR-WP-2023-038

Sub-national disparities in the global mobility of academic talent

Aliakbar Akbaritabar I akbaritabar@demogr.mpg.de
Maciej J. Dańko I danko@demogr.mpg.de
Xinyi Zhao I zhao@demogr.mpg.de
Emilio Zagheni I office-zagheni@demogr.mpg.de

Sub-national disparities in the global mobility of academic talent

Aliakbar Akbaritabar[1*], Maciej J. Dańko[1], Xinyi Zhao[1,2], and Emilio Zagheni[1]

[1]Max Planck Institute for Demographic Research, Rostock, Germany.
[2]Leverhulme Centre for Demographic Science, Department of Sociology, University of Oxford, Oxford, UK.
([*]Corresponding author, Email: akbaritabar@demogr.mpg.de)

Current version: September 8, 2023

Abstract

The migration of scholars has been often studied across countries, however, these studies have rarely focused on sub-national regions. We used data on 28+ million Scopus publications of 8+ million unique authors and geo-coded the affiliation addresses. Our results show that by focusing on the sub-national regions, the share of mobile scholars increases from 8% to 12.4%. We found that in all continents when a sub-national region is attractive for international migrants, it is also attractive for internal ones. The reverse is not true, though. For most continents, a depopulation is happening where scholars move abroad and their position is filled by scholars arriving from other sub-national regions inside the country. In the US, as an example, states in the mid-eastern area have the highest net rate of scholars leaving for other destinations inside the US, mostly on the west coast. In Europe, multiple countries show a similar trend that more developed provinces receive scholars from internal origins and send scholars to international destinations. Our results have implications for the global circulation of academic talent by adding more nuance to the generally accepted image of brain drain and brain gain. We highlight the interrelation between internal and international migration, specifically for regions constantly losing their academic workforce.

Keywords: Internal Migration; International Migration; Worldwide; Scientific mobility; Bibliometric data

Teaser

While only 8% of Scopus-published scholars experience international migration, 12.4% of them move between sub-national regions of countries.

1 Introduction

International mobility is known to favor career advancement and professional recognition of academic scientists (*1*). Among others, it facilitates the recombination of ideas (*2*, *3*) and the expansion of networks of collaborators, which, in turn, lead to better science and higher visibility of scholars (*4*). The benefits of geographic mobility are not felt only by individual scientists. Migrant scholars tend to be a highly selected group of researchers who are particularly productive and creative to start with (*5*). They favor knowledge circulation and are an engine of growth, especially in the destination countries (*2*, *3*). As a result, they are often the target of national policies (*6*) to attract 'the best and the brightest' (*7*).

The research on the determinants and consequences of migration of scholars has focused on international relocations (*8*, *9*). However, internal migration is generally expected to be higher than international migration and could have a bigger impact on the dynamics of populations of scholars across sub-national regions, with consequences in terms of the vibrancy of regions and their potential for serving as hubs of discovery. Despite the importance of internal migration for the scientific vitality of regions, our knowledge of patterns of migration of scholars at the sub-national level is extremely limited, and we have virtually no empirical evidence on how internal migration is inter-related with broader patterns of international migration. While we expect that the systems of internal and international migration are inter-connected (*10–13*), understanding their dynamics has been hindered by lack of appropriate data across different groups of migrants (*14–17*).

In the context of migration of scholars, bibliometric data offer previously unavailable opportunities (*8*, *9*) for jointly assessing the dynamics of internal and international migration processes, which typically cannot be assessed for other groups of migrants, including other types of high-skilled migrants. In this article, we propose an integrated framework to address the gap in understanding internal and international migration systems simultaneously. Among others, we quantify the extent to which *academic* talent circulation has happened "within" the national borders versus "between" countries; which sub-national regions are particularly attractive for scholars who move internally or internationally, and which ones have been losing scientists; how the dynamics of internal and international migration systems are inter-related. We leveraged large-scale bibliometric data comprising 28+ million publications by 8+ million unique authors, indexed in Scopus from 1996 to 2020, together with disambiguated and geocoded affiliation addresses. We prepared a novel and global database of migration of scholars covering both internal and international migration at GeoNames Admin-1 level, which enables us to assess differences between regions and countries from a comparative perspective.

2 Results

A new global database of internal and international migration of scientists. Fig. 1 shows an illustrative example of statistics produced as part of the new global database. In particular, it provides net migration rates (NMR) per 1,000 scholars, for the period from 2012-2017 to highlight the latest trends as a recent example from the available wider period. The rates shown are the difference between the sum of scholars entering a region during the six years considered from both internal and international origins and those exiting, divided by the sum of the population of scholars in the region. The majority of scholars in our database, do not experience mobility which is in line with the literature (*8*, *9*, *18–21*). However, focusing on the sub-national level instead of countries reduces the share of non-mobile scholars from 92% to 88% (based on mode-based strict measures, see methods section for a description), which highlights the higher prevalence of internal migration of scholars in contrast to international moves which has traditionally been the focus of the literature.

Fig. 1 shows that there is more nuance to the generally accepted image that some countries such as the US act as magnets attracting scholars in all states. Panel A shows that the US states are receiving a different rate of scholars. For instance, the balance of scholars sent or received in some states is negative indicated with red colors. This entails that the number of scholars exiting is higher than those entering these states. In addition, this figure shows that the general picture depicted with terms such as *brain drain* that some countries constantly lose their scholars is more nuanced. For instance, India at the country level has a negative NMR rate indicating sending more scholars than receiving but zooming in on sub-national regions (panel A) shows that some regions in India such as central and southern ones have a positive NMR. Since NMR uses the population of scholars as the denominator and could in principle be prone to under- or over-counting of scholars due to left or right censoring issues in the data (see methods section for further details), we use an additional measure, Migration Effectiveness Index (MEI (*22*, *23*)) that uses the absolute difference (nominator) and sum of scholars entering a region or exiting it (denominator).

The classification based on MEI is shown in Fig. S4 of the Supplementary Information, where we present analogous maps resembling those in Fig. 1.

Additionally, Fig. S5 in the Supplementary Information shows the NMR separated over internal and international rates. It shows that the trend of international migration does not necessarily match that of internal migration. Sub-national regions have different attractiveness in internal and international migration. For example, while the west coast in the US is an attractive region for both internal and international scholars, it is not the same case for middle and eastern states. Eastern states are more attractive for international scholars than internal ones. In general, many of the states in the mid-eastern area of the US are sending states and have a negative net migration rate (indicated with red colors), and western regions are receiving (green colors). This stark difference between internal and international migration is also observed in provinces in Australia, South America, India, Turkey, Saudi Arabia, Iran, and China. Zooming in on Europe (panel B in Fig. S5) shows even more evident differences. While many provinces in Germany, Italy, France, Spain, the UK, and the Netherlands are indicated with red colors and send scholars to international destinations in panel A, there is a large difference between internal migration to these provinces in panel B. For instance, the Northern and middle provinces in Italy, southern regions in France, England in the UK, and some regions in Spain are receiving scholars from other sub-national origins inside these countries. This could be a signal of *academic depopulation* in these regions where academic positions of those leaving for international destinations are filled by internal migrants.

Global trends of temporal change in internal and international migration to and from sub-national regions. Fig. 2 compares the range of differences and inequality of attractiveness between sub-national regions at the continent level using the Gini coefficient over time for four types of migration (i.e., internal, international, in- and out-migration). It shows that population weighting lowers Gini coefficients in all continents. This is expected since small regions will introduce more sampling variation and weighted averages might correct this problem. In Europe, weighting increases the spread in inequality of rates between internal and international migration, however, it narrowed this spread in North America. In Asia, and to a lesser extent South America, we are seeing a pattern of increasing inequalities in international migration and decreasing inequalities in internal migration. Africa shows a steady reduction in inequality across all types of migration. Europe, North America, and Oceania show a rather steady trend in inequality across all types of migration over time. This means, there are stable flows of incoming or outgoing migrant scholars in these continents that have not changed much during our observation period. On the contrary, Asia and South America present a change in their Gini coefficient over time indicating that they have gone through an era of transformation in scholarly migration. This could indicate that some regions have changed and become popular sending or receiving regions over time.

Fig. 3A and 3B provide a more nuanced temporal picture per sub-national region by showing the slopes of quasi-Poisson regressions estimating the temporal trends from 1998 to 2017 for in- and out-migration rates. We include all global sub-national regions having more than 25 scholars in at least one observation year, additionally a minimum of at least 10 observation years. Fig. 3A shows that in all continents, the lower left quadrant, indicated with a green shade, contains the highest percentage of regions. It means that both internal and international in-migration decreases over time and has negative slopes. Looking at internal and international migration separately, most regions are in the lower and left quadrants, respectively. It indicates decreasing trends for both types of migration. In all continents except Europe and North America, we observe that as international in-migration increases, internal in-migration increases. This means, if a region is attractive for hosting researchers, it receives scholars from both internal and international origins. Europe and North America have high p-values which suggest that the observed trend may not be reliable. Fig. 3B shows that Africa, Asia and South America have the highest percentage of regions located in the upper quadrants on the Y-axis and have increasing trends in internal out-migration. This means that sending scholars to other regions within a given country increases over time. In Europe, North America, and Oceania, the largest proportion of regions are in the bottom quadrants. It means that in most regions, out-migration

decreases over time.

The relationship between internal and international migration worldwide. Fig. 4 shows that in all continents, most regions are in the upper right quadrant, indicated with a green shade, which determines the positive correlation between internal and international migration for both out-migration and in-migration. A positive correlation means that either both internal and international trends are increasing over time, or both trends are decreasing. That is, if a region receives scholars, it consistently receives scholars from both domestic and foreign origins, and if a region sends scholars, it consistently sends scholars to both domestic and international destinations. This further confirms the trends shown in Fig. 3A and 3B.

3 Discussion

We prepared a global database of migration of scholars at the level of internal migration within countries and at the level of international migration. This database enabled us to address a documented gap in the literature to study the two interrelated migration systems (*12*, *13*) in a unified framework (*10*, *11*). On a global scale, we developed indexes of internal and international migration of scholars (*22*, *23*).

A clear pattern is found in in-migration rates for all continents. An attractive region, receiving scholars from international origins is also attractive in receiving scholars from internal origins. The reverse, being a popular sending origin to internal and international destinations is only true for African and European regions. That means sub-national regions in Africa and Europe send scholars to both types of destinations and this trend has been generally increasing over the 24 years studied. In Asia, North America, South America, and Oceania the opposite trend is observed which could indicate an academic depopulation effect, where regions sending scholars to international destinations, in turn, receive scholars from internal origins.

We found a group of countries where internal scholarly migration was more prevalent and effective. These countries with larger science systems can hire their graduates and promote them to permanent positions (*24*, *25*). Internal scholarly migration is more prevalent and has a higher impact on these science systems. Our more detailed investigation of internal and international scholarly migration in sub-national regions of these countries showed a highly dynamic trend. Some of the less developed regions were mainly sending regions whose talent is hired elsewhere inside the country or in international destinations.

The dynamics of talent circulation between sub-national regions of the countries were insightful. For specific countries (e.g., Italy), we observed a higher rate of incoming scholars in more developed regions and the leave of talent from the less developed regions to both internal and international destinations. While specific countries such as the US act as an international magnet of academic talent at the country level, at the sub-national level, some states (e.g., mid-eastern ones) were losing scholars to other states (western ones) and go through an academic depopulation process, similar to observed trends for provinces in Europe. This was possible to observe only after zooming in on the sub-national level and considering the inter-relationship between internal and international migration trends which is a contribution of our curated and novel database.

To conclude, based on our results, our methodology in re-purposing bibliometric data offers unprecedented opportunities to study *internal* and *international* scholarly mobility and systems of migration (*10*, *11*) in an integrated framework. The constructed macro global database enables us to identify attractive destinations of scholarly mobility for national and international academics. It enables us to see which regions are traditionally sending regions and which ones are the receiving regions. In addition, it allows further investigation of potential underlying factors (e.g., development or lack thereof) driving the observed trends (*8*). This type of investigation is influential in the area of *high-skilled migration* and *global talent circulation* (*1*, *4*, *26*, *27*). Furthermore, there are gender (*9*, *28*) and disciplinary differences (*29*) in the observed trends that need further investigation. The meaning and implications of higher and lower scholarly mobility on circulation of ideas in a country and recombination of ideas (*30*–*33*) need

further probes. We could ask if there is a type of gravity (*21*) of scholars (those more prominent) or institutions (those with higher prestige as shown by Clauset *et al.* (*34*)) attracting the brain and driving brain circulation patterns?

Our study and database have certain limitations. Academic entities' names (e.g., author and organization names) need disambiguation and cleaning (*35–38*). We have used the most reliable available database which is proprietary (*39*) and have done progress in that direction, but the used methods are still prone to error (e.g., see a discussion in *38*). In addition, higher-level epistemic questions should be answered while re-purposing bibliometric data for migration research (*33*, *40–42*), for instance, assigning the country of affiliation in the first publication as the country of origin for academic mobility is prone to error since that could simply be the country of graduation (*8*), hence, it should only be considered as the academic origin and not a proxy for nationality. There is a publication delay (*43*) that can hinder proper identification of the mobility period. Furthermore, these data are limited to only those scholars who have actively published in scholarly journals. These scholarly journals are limited to the ones indexed by the bibliometric database used and in most cases are dominated by English-speaking publications with lower coverage of non-English ones. Further, some scholars and disciplines are not publishing as often as hard sciences and our results are driven by scholars and disciplines who tend to publish more (*44*).

We curated the global database of scholarly mobility and answered some previously non-addressable questions which were due to a lack of data on internal and international migration (*10*, *45*). Our database enables addressing such questions for the subset of highly skilled migrants, i.e., scholars. Our results have certain implications for theories of migration by showing that different sub-national regions in a country with specific development history could exhibit specific migration trajectories (*8*). Since the links between the migration of scholars, knowledge diffusion and collaboration are empirically and theoretically understudied (*21*), our study highlights that such links need to be studied both within countries and in the case of international migration of scholars. In addition, the case of scholars who would like to migrate but due to different reasons could not do so and its implications for their scientific work should be studied similarly to the case of the general population's involuntary immobility discussed by Carling (*46*). As shown previously by us in Sanliturk *et al.* (*8*), different subsets of the population could be affected by different pull and push factors. Scientists' migration behavior with development might be exceptional and different from the general population (*21*). Future research could include investigating more substantive questions. It is beneficial to investigate the intertwined ties between scholarly *mobility* and *collaboration* and to better identify the causal direction between the two, at the micro (individual scholar) level which is still missing from the literature (*27*). It is necessary to determine whether the causation is from mobility to collaboration (*47*) or bidirectional (*18*, *48*) or inconclusive based on the vast literature advocating one way or the other (*49*).

Acknowledgements

We would like to thank our colleague Tom Theile for technical support in preparing the raw bibliometric data, Chiara Franzoni, Benedetto Lepori and attendees of RISIS research seminars for helpful comments on an earlier version of our results.

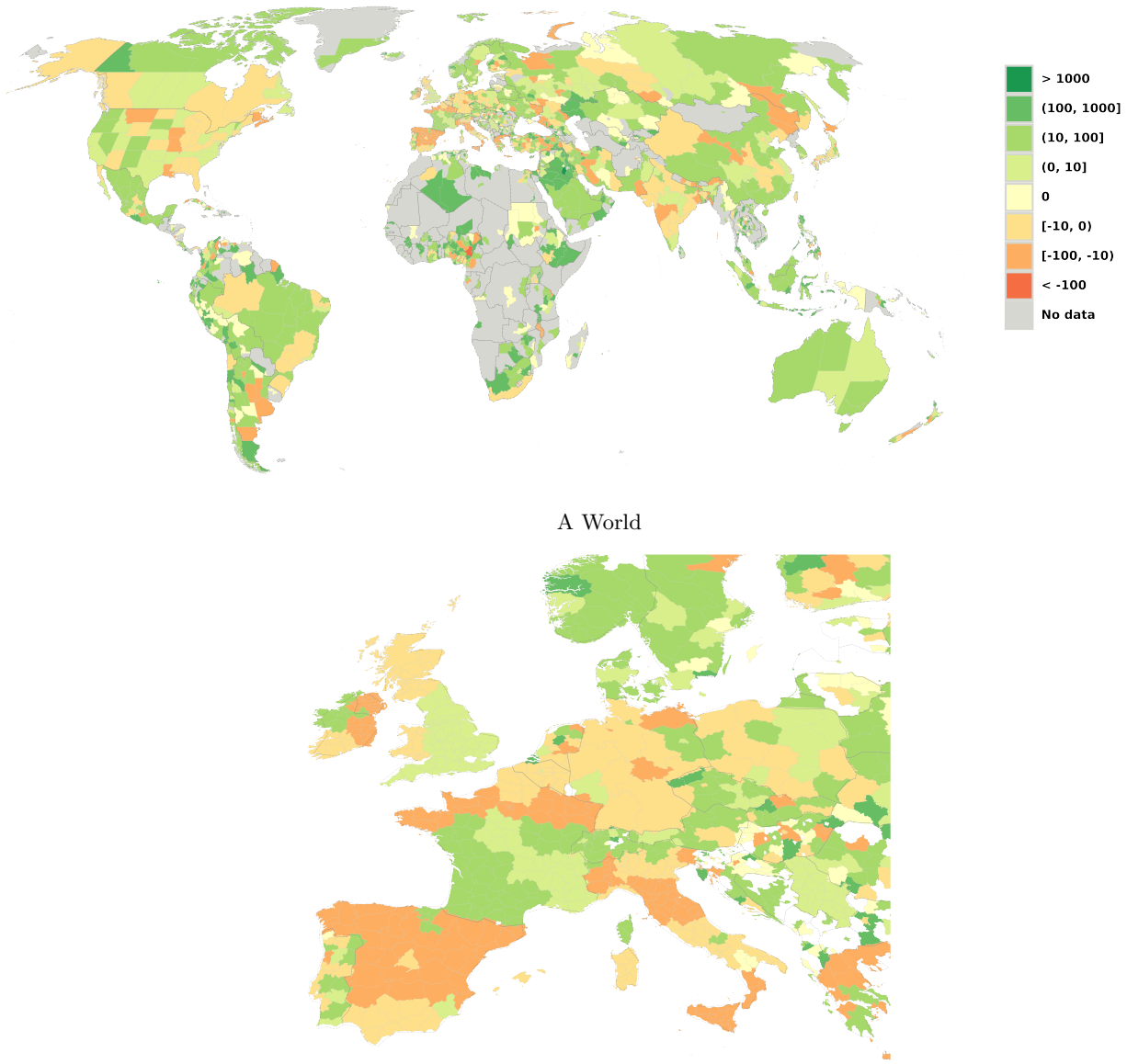
Funding: This study has received access to the bibliometric data through the project "Kompetenzzentrum Bibliometrie" (*50*) and we acknowledge their funder Bundesministerium für Bildung und Forschung (grant number 16WIK2101A).

Authors' contributions: Conceptualization: A.A., E.Z.; Data Curation: A.A., X.Z., M.J.D.; Methodology: A.A., M.J.D.; Formal analysis: M.J.D.; Visualization: M.J.D., A.A.; Software: M.J.D., A.A., X.Z; Writing - original draft: A.A., E.Z., Writing - review & editing: A.A., M.J.D., E.Z., X.Z.

Competing interests: The authors declare that they have no competing interests.

Availability of data and materials: We publicly share aggregate sub-national region level data and scripts

that allow reconstructing our analysis in this GitHub repository (**will be added upon acceptance; anonymized link is available for peer review purposes**).



A World

B Europe

Fig. 1: **Combined internal and international net migration rate per 1000 scholars at the province level worldwide (A) and Europe (B)**. Colors present the net migration rate of scholars sent (negative, red colors) or received (positive, green colors) in a year for the latest six years as an illustrative example of the 25 years studied, and yellow shows a balanced flow. Color scales are kept similar in the two maps to allow comparison while the upper and lower class includes above 1000 and below -100 net rates, respectively. Numbers printed on the legend are the combined net rate of scholars sent or received to/from international and internal origins/destinations showing which regions are sending more scholars than they receive or the reverse.

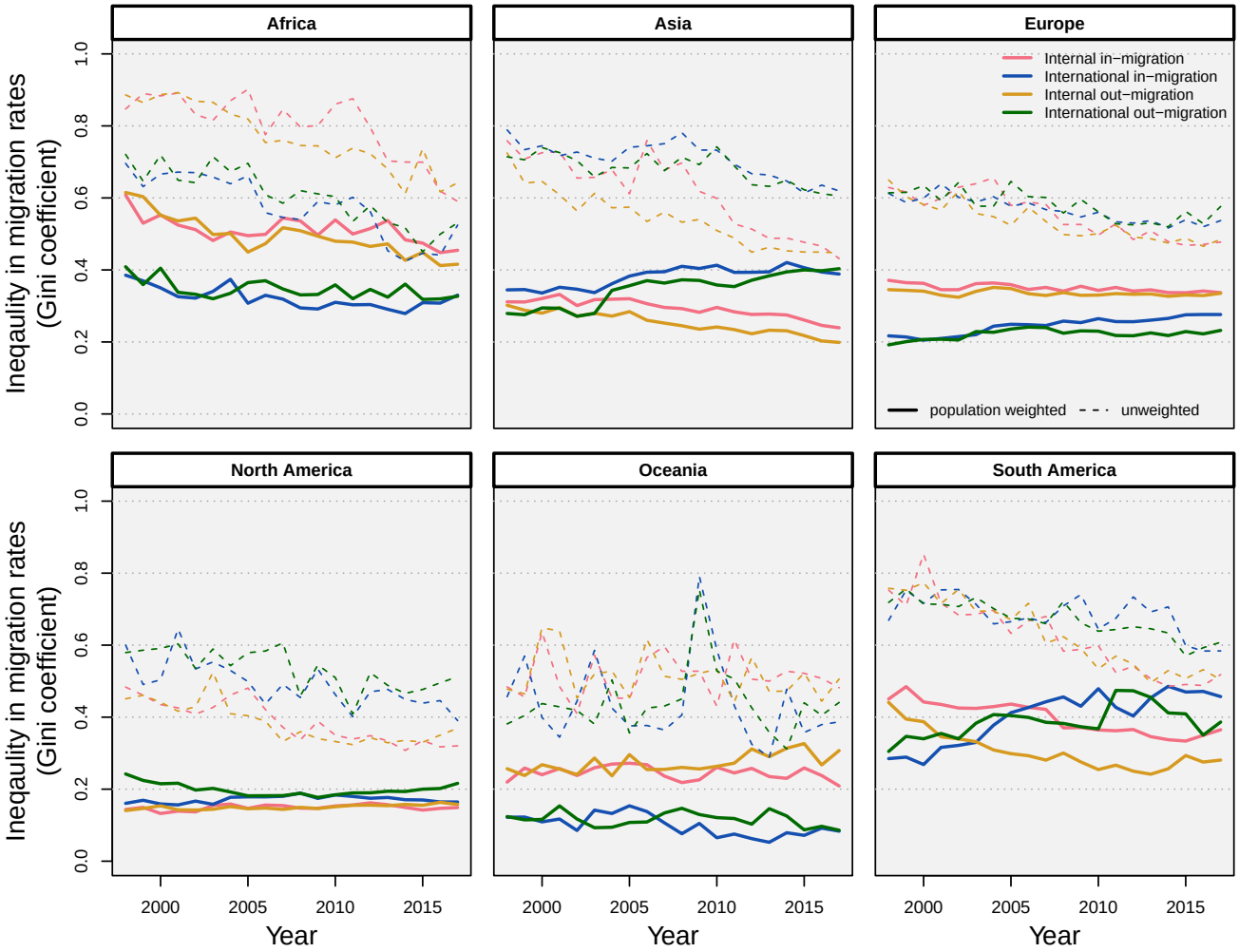
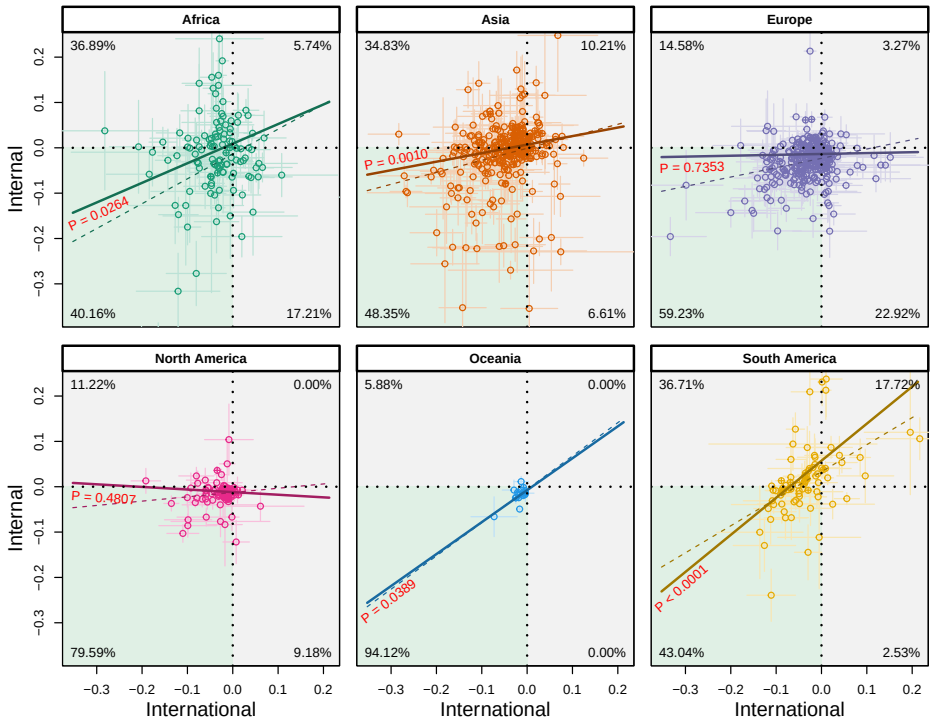
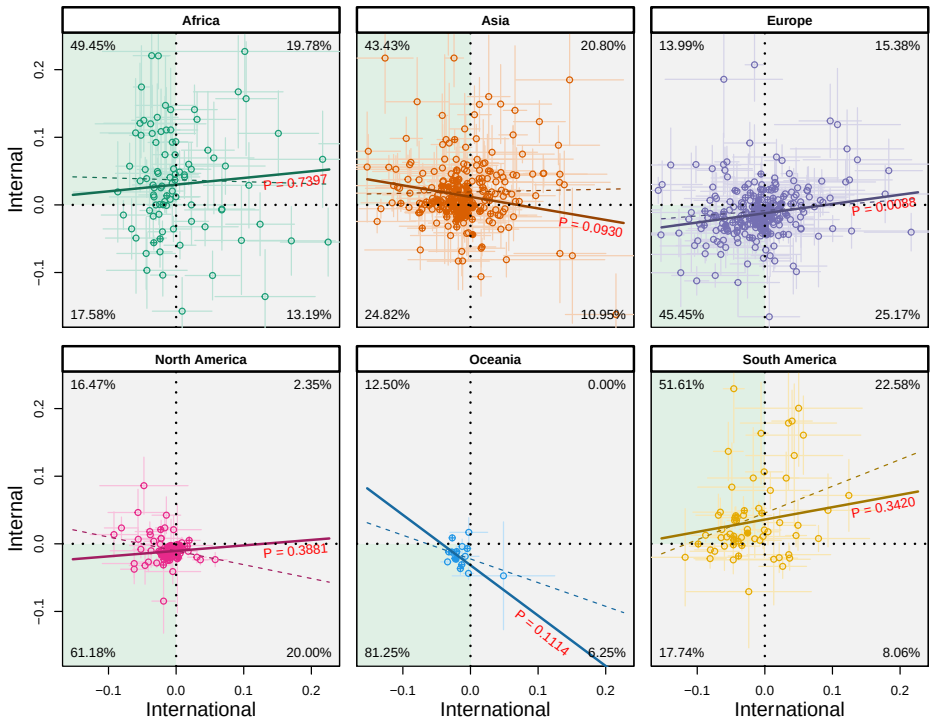


Fig. 2: **Year-specific region-based Gini coefficients for different continents, and migration types.** Two versions are considered: (1) unweighted, dashed lines - small regions are treated the same as large regions, but their estimates may be biased (high uncertainty when there are a small number of scholars), and (2) weighted, solid lines - population size is used as a weight to control for the bias resulting from regions with a very small number of scholars and large sampling variation in rates, however, regions with a high number of scholars may dominate the estimated inequity measure.



A In-migration



B Out-migration

Fig. 3: **Slopes of region-specific time trends of log-in-migration (A) and log-out-migration (B) rates obtained from multiple independent quasi-Poisson regressions.** Circles represent sub-national regions where filled circles have a larger than 10,000 population of scholars; semi-filled circles: between 1000 and 10,000; open circles: equal to or lower than 1000. Percentages printed in each corner denote the fraction of regions per quadrant. Horizontal and vertical lines denote ± standard errors of each regional estimate. The lines represent fitted models; thin dashed line: standard OLS; thick solid line: OLS accounting for heteroscedasticity problems in dependent and independent variables. P-values are related to slopes of the latter model.

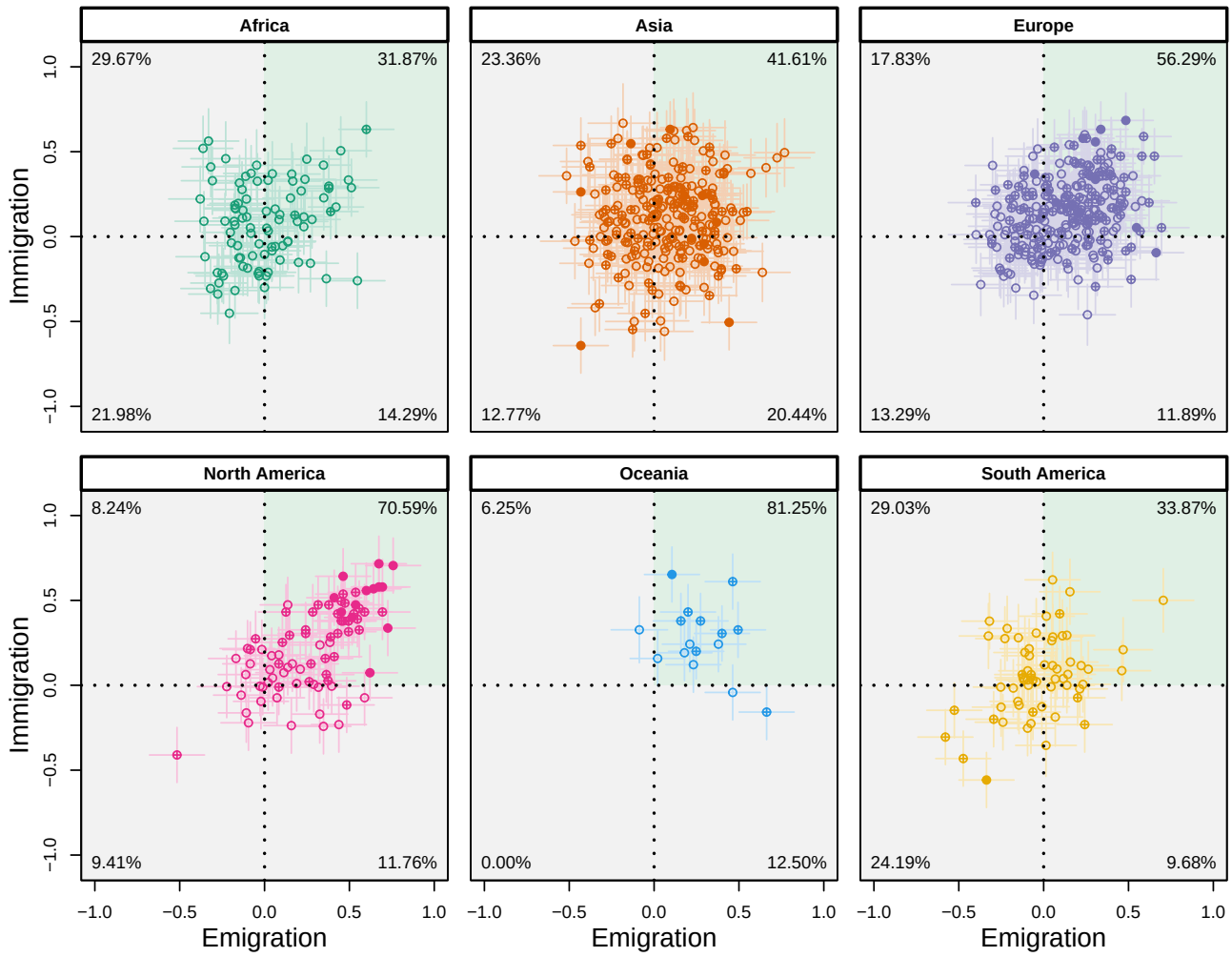


Fig. 4: **Kendall rank correlation coefficients between internal and international time trends for different sub-national regions (circles) investigating simultaneously the inter-relationships between out-migration and in-migration.** Circles represent sub-national regions where filled circles have a larger than 10,000 population of scholars; semi-filled circles: between 1000 and 10,000; open circles: equal to or lower than 1000. Percentages printed in each corner denote the fraction of regions per quadrant. Upper-right quadrant: the correlation is positive for both in-migration and out-migration. Bottom-right quadrant: the correlation is negative for both in-migration and out-migration. Similar logic applies for the remaining quadrants, however, the positive correlation between internal and international is recorded only for one data type. The positive correlation between internal and international means that for both data types, there are either positive or negative time trends. The negative correlation means that for both data types, we see opposite directions in internal and international migration.

References

1. G. Scellato, C. Franzoni, P. Stephan, en, *Science* **356**, Publisher: American Association for the Advancement of Science Section: Policy Forum, 694–694, ISSN: 0036-8075, 1095-9203, (2021; https://science.sciencemag.org/content/356/6339/694.2) (May 2017).
2. R. Freeman, *Science* **356**, Publisher: American Association for the Advancement of Science, 696–697, (2022; https://www.science.org/doi/full/10.1126/science.aan5197) (May 2017).
3. V. Aman, *Quantitative Science Studies*, Publisher: MIT Press, 1–17, (2020; https://doi.org/10.1162/qss_a_00028) (Mar. 2020).
4. C. R. Sugimoto *et al.*, *Nature* **550**, 29–31, ISSN: 0028-0836, 1476-4687, (2020; http://www.nature.com/articles/550029a) (Oct. 2017).
5. R. Veugelers, *Science* **356**, Publisher: American Association for the Advancement of Science, 695–696, (2022; https://www.science.org/doi/full/10.1126/science.aan3920) (May 2017).
6. OECD, *The Global Competition for Talent Mobility of the Highly Skilled: Mobility of the Highly Skilled*, en (OECD Publishing, Sept. 2008), ISBN: 978-92-64-04775-4, (https://read.oecd-ilibrary.org/employment/the-global-competition-for-talent_9789264047754-en#page1).
7. D. Kapur, J. McHale, *Give Us Your Best and Brightest: The Global Hunt for Talent and Its Impact on the Developing World*, en (Center for Global Development, 2005), ISBN: 978-1-933286-03-7.
8. E. Sanliturk, E. Zagheni, M. J. Dańko, T. Theile, A. Akbaritabar, *Proceedings of the National Academy of Sciences* **120**, e2217937120, (2023; https://www.pnas.org/doi/10.1073/pnas.2217937120) (Jan. 2023).
9. X. Zhao, A. Akbaritabar, R. Kashyap, E. Zagheni, *Proceedings of the National Academy of Sciences* **120**, e2214664120, (2023; https://www.pnas.org/doi/10.1073/pnas.2214664120) (Mar. 2023).
10. R. Skeldon, en, *Population, Space and Place* **12**, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/psp.385, 15–30, ISSN: 1544-8452, (2021; https://onlinelibrary.wiley.com/doi/abs/10.1002/psp.385) (2006).
11. R. King, R. Skeldon, en, *Journal of Ethnic and Migration Studies* **36**, 1619–1646, ISSN: 1369-183X, 1469-9451, (2021; http://www.tandfonline.com/doi/abs/10.1080/1369183X.2010.489380) (2010).
12. R. J. Pryor, *International Migration Review* **15**, Publisher: SAGE Publications Inc, 110–129, ISSN: 0197-9183, (2021; https://doi.org/10.1177/019791838101501s08) (1981).
13. M. Ellis, en, *Population, Space and Place* **18**, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/psp.666, 196–208, ISSN: 1544-8452, (2022; https://onlinelibrary.wiley.com/doi/abs/10.1002/psp.666) (2012).
14. A. Adepoju, *International Social Science Journal* **50**, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-2451.00151, 387–395, ISSN: 1468-2451, (2021; https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-2451.00151) (1998).
15. A. Bernard, F. Perales, *Journal of Ethnic and Migration Studies*, 1–21, ISSN: 1369-183X, 1469-9451, (2021; https://www.tandfonline.com/doi/full/10.1080/1369183X.2020.1868983) (2021).
16. G. J. Hugo, *Population, Space and Place* **22**, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/psp.1930, 651–668, ISSN: 1544-8452, (2021; https://onlinelibrary.wiley.com/doi/abs/10.1002/psp.1930) (2016).
17. E. Zagheni, V. R. K. Garimella, I. Weber, B. State, presented at the Proceedings of the 23rd International Conference on World Wide Web, pp. 439–444.

18. H. D. Boekhout, V. A. Traag, F. W. Takes, en, *Network Science* **9**, Publisher: Cambridge University Press, 354–386, ISSN: 2050-1242, 2050-1250, (2021; https://www.cambridge.org/core/journals/network-science/article/investigating-scientific-mobility-in-coauthorship-networks-using-multilayer-temporal-motifs/4A8730DC440D7BE7EF4E6306AEE6ACBD) (2021).

19. F. Momeni, F. Karimi, P. Mayr, I. Peters, S. Dietze, *Journal of Informetrics* **16**, 101280, ISSN: 1751-1577, (2022; https://www.sciencedirect.com/science/article/pii/S1751157722000323) (2022).

20. V. Macháček, M. Srholec, M. R. Ferreira, N. Robinson-Garcia, R. Costas, *Science and Public Policy*, ISSN: 0302-3427, (2021; https://doi.org/10.1093/scipol/scab064) (Sept. 2021).

21. M. Czaika, S. Orazbayev, en, *Applied Geography* **96**, 1–10, ISSN: 0143-6228, (2021; https://www.sciencedirect.com/science/article/pii/S0143622817305428) (July 2018).

22. M. Bell *et al.*, en, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **165**, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-985X.t01-1-00247, 435–464, ISSN: 1467-985X, (2021; https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-985X.t01-1-00247) (2002).

23. A. Miranda-González, S. Aref, T. Theile, E. Zagheni, en, *EPJ Data Science* **9**, Number: 1 Publisher: Springer Berlin Heidelberg, 34, ISSN: 2193-1127, (2021; https://epjds.epj.org/articles/epjdata/abs/2020/01/13688_2020_Article_252/13688_2020_Article_252.html) (2020).

24. J. Grilli, S. Allesina, en, *Proceedings of the National Academy of Sciences* **114**, 7600–7605, ISSN: 0027-8424, 1091-6490, (2020; http://www.pnas.org/lookup/doi/10.1073/pnas.1703513114) (July 2017).

25. S. Allesina, en, *PLoS ONE* **6**, ed. by E. von Elm, e21160, ISSN: 1932-6203, (2020; https://dx.plos.org/10.1371/journal.pone.0021160) (Aug. 2011).

26. T. Wiesel, en, *Nature* **510**, Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7504 Primary_atype: Comments & Opinion Publisher: Nature Publishing Group Subject_term: Developing world;Institutions;Resea management Subject_term_id: developing-world;institutions;research-management, 213–214, ISSN: 1476-4687, (2021; https://www.nature.com/articles/510213a) (June 2014).

27. S. Appelt, B. van Beuzekom, F. Galindo-Rueda, R. de Pinho, en, in *Global Mobility of Research Scientists*, ed. by A. Geuna (Academic Press, San Diego, Jan. 2015), pp. 177–213, ISBN: 978-0-12-801396-0, (2022; https://www.sciencedirect.com/science/article/pii/B9780128013960000077).

28. X. Zhao, S. Aref, E. Zagheni, G. Stecklov, *Scientometrics*, ISSN: 1588-2861, (2022; https://doi.org/10.1007/s11192-022-04351-4) (2022).

29. G. Laudel, J. Bielick, en, *Research Policy* **48**, 103800, ISSN: 0048-7333, (2021; https://www.sciencedirect.com/science/article/pii/S0048733319301209) (Nov. 2019).

30. G. Scellato, C. Franzoni, P. Stephan, en, *Research Policy* **44**, 108–120, ISSN: 0048-7333, (2020; http://www.sciencedirect.com/science/article/pii/S0048733314001358) (Feb. 2015).

31. S. Sbalchiero, A. Tuzzi, en, *International Migration* **55**, 171–187, ISSN: 0020-7985, 1468-2435, (2021; https://onlinelibrary.wiley.com/doi/10.1111/imig.12340) (Aug. 2017).

32. C. Saint-Blancat, en, *Higher Education Policy* **31**, 37–54, ISSN: 1740-3863, (2022; https://doi.org/10.1057/s41307-017-0042-z) (Mar. 2018).

33. G. Laudel, *Scientometrics* **57**, 215–237, ISSN: 1588-2861, (2021; https://doi.org/10.1023/A:1024137718393) (2003).

34. A. Clauset, S. Arbesman, D. B. Larremore, *Science Advances* **1**, e1400005, ISSN: 2375-2548, (2020; https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.1400005) (2015).

35. J. Wu, X.-H. Ding, *Scientometrics* **96**, 683–697, ISSN: 0138-9130, 1588-2861, (2019; http://link.springer.com/10.1007/s11192-013-0978-8) (2013).

36. A. Tekles, L. Bornmann, *Quantitative Science Studies*, Publisher: MIT Press, 1–38, (2020; https://doi.org/10.1162/qss_a_00081) (2020).

37. P. Donner, C. Rimmert, N. J. van Eck, *Quantitative Science Studies* **1**, 150–170, (2020; https://doi.org/10.1162/qss_a_00013) (Dec. 2019).

38. A. Akbaritabar, *Quantitative Science Studies* **2**, 753–777, ISSN: 2641-3337, (2021; https://doi.org/10.1162/qss_a_00131) (2021).

39. J. Baas, M. Schotten, A. Plume, G. Côté, R. Karimi, *Quantitative Science Studies* **1**, 377–386, ISSN: 2641-3337, (2021; https://doi.org/10.1162/qss_a_00019) (2020).

40. V. Aman, *Scientometrics* **117**, 227–247, ISSN: 1588-2861, (2020; https://doi.org/10.1007/s11192-018-2864-x) (2018).

41. H. F. Moed, M. Aisati, A. Plume, *Scientometrics* **94**, 929–942, ISSN: 0138-9130, 1588-2861, (2020; http://link.springer.com/10.1007/s11192-012-0783-9) (2013).

42. H. F. Moed, G. Halevi, *Scientometrics* **101**, 1987–2001, ISSN: 1588-2861, (2021; https://doi.org/10.1007/s11192-014-1307-6) (2014).

43. B.-C. Björk, D. Solomon, en, *Journal of Informetrics* **7**, 914–923, ISSN: 1751-1577, (2022; https://www.sciencedirect.com/science/article/pii/S1751157713000734) (2013).

44. A. J. Lotka, *Journal of the Washington Academy of Sciences* **16**, Publisher: Washington Academy of Sciences, 317–323, ISSN: 0043-0439, (2022; https://www.jstor.org/stable/24529203) (1926).

45. F. Willekens, D. Massey, J. Raymer, C. Beauchemin, *Science* **352**, Publisher: American Association for the Advancement of Science, 897–899, (2022; https://www.science.org/doi/10.1126/science.aaf6545) (May 2016).

46. J. Carling, *Journal of Ethnic and Migration Studies* **28**, Publisher: Routledge _eprint: https://doi.org/10.1080/1369183012015–42, ISSN: 1369-183X, (2023; https://doi.org/10.1080/13691830120103912) (Jan. 2002).

47. M. Kato, A. Ando, en, *Scientometrics* **110**, 673–694, ISSN: 1588-2861, (2021; https://doi.org/10.1007/s11192-016-2183-z) (2017).

48. Z. Chinchilla-Rodríguez *et al.*, English, *Frontiers in Research Metrics and Analytics* **3**, Publisher: Frontiers, ISSN: 2504-0537, (2020; https://www.frontiersin.org/articles/10.3389/frma.2018.00017/full) (2018).

49. N. Netz, S. Hampel, V. Aman, en, *Research Evaluation* **29**, Publisher: Oxford Academic, 327–351, ISSN: 0958-2029, (2020; https://academic.oup.com/rev/article/29/3/327/5879506) (2020).

50. German Competence Network for Bibliometrics, *Kompetenzzentrum Bibliometrie (KB)*, 2021, (2021; https://bibliometrie.info/).

Supplementary Materials for
Sub-national disparities in the global mobility of academic talent

Authors: Aliakbar Akbaritabar*, Maciej J. Dańko, Xinyi Zhao, and Emilio Zagheni.

*Corresponding author, Email: akbaritabar@demogr.mpg.de

This PDF file includes:

1. Materials and Methods
2. Additional detail on data and pre-processing steps
3. Additional measures
4. Net internal and international scholarly migration rates for the US
5. More on Aggregated Net Migration Rate, and Migration Effectiveness Index measures
6. Figures S1-S7

1 Materials and Methods

We use publications data from Elsevier's 2020 snapshot of Scopus that is provided to us by the German Competence Network for Bibliometrics (*1*) through the Max Planck Digital Library. We limit the data to only "*article*" and "*review*" publications to have the highest quality of metadata. After disambiguation and geo-coding to the sub-national level, following methodology outlined by us previously in Akbaritabar (*2*) using a local installation of Research Organization Registry (ROR) Application Programming Interface (API), we exclude publications without usable addresses. Our data-set includes 28,461,324 publications from 1996 to 2020 by a total of 8,225,368 Scopus published and disambiguated authors. Note that in our statistical models, we include the data from 1998 to 2017 to reduce left- and right- censoring effects in the data.

We use a strict measure based on the mode-region of affiliation to identify mobility events which is prone to less noise and fluctuation based on our previously published investigation and analysis (e.g., see *3–5*). We considered all affiliation countries of a Scopus author ID in a single year. If there is more than one country in a year, we take the *mode* (the most frequent affiliation) of all countries and in the case of multiple modes, we choose the one that was present in the closest previous years. When all mode countries are unique and new, we choose one randomly (*3*, *6*). A migration event is recorded when the mode of the country of residence in year t changes in year of next observation. The same happens for the sub-national regions and the residence is the mode region in a given year. Furthermore, we assume a *two years* preparation time for all publications to cover disciplinary differences in publication delay (*7*). If there are gaps in publication years (e.g., authors are not publishing continuously), we backward fill each publication year for two years and assume the author's residence to have been changed two years earlier (*6*). If there is enough evidence (i.e., continuous publication activity), we consider the year a modal affiliation changes, as the migration year. Note that in literature, migration is considered a longer-term mobility event (*8*), and for shorter-term moves, labels such as *workplace mobility* have been used (*9*). Here by the mode-based definition, mobility needs to happen over a minimum of two years (and cannot happen in the same year), hence, we consider this a longer-term move than travelling or temporary stays which might not result in a change in academic affiliation address. Therefore, we use the words *mobility* and *migration* interchangeably throughout the text.

To aggregate the count of migration events to the region, country, and continent levels, we calculated different measures. We calculated the net migration rate as in equation 1 that accounts for the incoming and outgoing

population of scholars over the total number of scholars in a region or country in a given year and extended it to the most recent six-year period.

$$NMR_{i,t,k} = 1000 \times \frac{I_{i,t,k} - E_{i,t,k}}{N_{i,t}} \quad (1)$$

where i is the sub-national region. t is the year. Subscript k shows the type of data, i.e., internal or international. $I_{i,t}$ is the inflow of scholars entering a region and $E_{i,t}$ is the outflow of scholars exiting that region. $N_{i,t}$ is the total number of scholars in the region in a given year. We present results based on NMR of *internal* and *international* scholarly mobility to highlight the interdependence between these two systems of migration (e.g., as emphasized by Skeldon (*10*) and King and Skeldon (*11*)).

1.1 Statistical methods

Migration trends within continents (figure S1) were calculated using Generalized Additive Mixed Models (GAMM). GAMM are an extension of the Generalized Linear Mixed Models (GLMM) that are applied to model non-linear relationships using penalized smooth functions of selected predictor variables. We use a quasi-Poisson distribution (log as the link function) to account for potential over-dispersion and Restricted Maximum Likelihood (fREML) as a fitting method. In GAMM, migration rates were modeled using count data models with offset set to the log number of scholars in a specific region and for a given year (exposures approximation). The model was fitted using the following formula

$$log\ counts \sim s(year) + s_{fs}(year, region) + offset$$

where $s(year)$ is a smooth function of year using p-spline basis and $s_{fs}()$ is a factor smooth interaction used to model random effects. We also tested an alternative simpler model

$$log\ counts \sim s(year) + s_{re}(year, region) + s_{re}(region) + offset$$

where $s_{re}()$ is used to model simple uncorrelated random effects. Both models gave very similar results. Please see (*12*) for different application of similar models.

Slope coefficients shown in Fig. 3A and 3B in the main text were used to measure a general time trend and were estimated separately for each region. We employed count data models assuming migration counts as the dependent variable, with offset set to the log number of scholars in a specific region and for a given year (similarly to the approach used in GAMMs mentioned earlier). The quasi-Poisson distribution (with a log link function) was chosen because it can account for overdispersion, a common issue encountered when fitting standard Poisson models. The model was fitted in R (*13*) using the *gam*() function of the *mgcv* package (*14*). To quantify inequality in migration rates among regions we used Gini coefficient.

The Kendall rank coefficient (*15*) was used to measure the ordinal association between two variables and it is based on ranking the elements of the sample. In Fig. S2 and S3 we calculated Kendall rank coefficients independently for each regional trend. In Fig. 4 in the main text Kendall rank coefficients were also calculated for each region separately, but we use it to measure the association between internal and international migration rates. Migration rates were calculated as counts of migration events divided by the number of scholars.

Figures 3A and 3B in the main text, as well as S2 and S3 show two regression lines. Thin lines represent standard least squares regression (OLS). The thick lines show the OLS, which takes into account the heteroscedasticity problems associated with different standard errors of the dependent and independent variables. To correct for the heteroscedasticity of the dependent variable, we used the algorithm proposed by Lewis and Linzer (*16*). The heteroscedasticity of the independent variable was adjusted by assuming weights equal to the reciprocal of the variance of each observation.

Fig. S1 shows the temporal change in two systems of migration, i.e., internal and international, compared across in-migration and out-migration. In all continents, the in-migration for both international (blue line) and internal (red line) stays above the out-migration for international (green line) and internal (yellow line). In Africa and Oceania, the international rates have a large gap with the internal rates and the order of magnitude in internal is much higher, i.e., close to twice the international one. In Europe and North America, the reverse is happening and the internal rates are above the international rates while the order of magnitude is closer and the gap is less pronounced. In most continents, the trends are decreasing over time, meaning the migration of scholars is decreasing with the exception of internal out-migration in Africa, Asia and South America and internal in-migration in South America.

2 Supplementary methods and results

Here we present further information on our data, pre-processing steps needed, and additional and detailed results on specific country cases to complement the image presented in the main text.

2.1 Additional detail on data and pre-processing steps

We use publications data from Elsevier's 2020 snapshot of Scopus that is provided to us by the German Competence Network for Bibliometrics (*1*). It includes 28,461,324 "article" and "review" publications from 1996 to 2020. We limit the publication types to only these two to have the highest possible accuracy of metadata based on the evaluations by German Competence Network for Bibliometrics (*1*) and ourselves (*3*, *4*, *6*). Re-purposing publication data for our goals requires extensive data processing and cleaning. Further, bibliometric data needs fine-grained cleaning, encoding countries and geographical regions of affiliation. Academic organization and author names need to be disambiguated since for author names, homonyms and name changes occur and for organizations, spelling errors or use of different order of name parts happen (*17*). For author names, we use Scopus's author IDs with 98.3% precision (no publication by others are included in X's publication list) and 90.6% recall (all publications by X are included) (*18*). For the academic affiliations and organization names, we use the methodology outlined by Akbaritabar (*2*) which uses Research Organization Registry (ROR) application programming interface (API). By sending affiliation strings from Scopus to ROR API, we can identify similar affiliations with spelling and name order differences and group them under unique addresses. In addition, after processing bibliometric data, we complement them with GeoNames' codes for sub-national country regions in the highest granularity level (GeoNames Admin 1) which is roughly equivalent to NUTS 1 level in Europe and states and provinces in most other countries such as the US.

2.2 Additional Measures

In addition to the measures presented in the methods section, and to provide a more suitable measure of comparison between multiple countries, we calculated the Aggregated Net Migration Rate (ANMR) as in equation 2. To further complement our analysis and control the effectiveness of the migration events in redistributing the scholars' population between regions inside a country, we calculated Migration Effectiveness Index (MEI) (*5*, *19*) as in equation 3.

$$ANMR_t = 100 \times \frac{0.5 \sum_i |I_{i,t,k} - E_{i,t,k}|}{\sum_i N_{i,t}} \tag{2}$$

$$MEI_t = 100 \times \frac{\sum_i |I_{i,t,k} - E_{i,t,k}|}{\sum_i (I_{i,t,k} + E_{i,t,k})} \tag{3}$$

The main difference between ANMR and MEI (as described in Miranda-González *et al.* (*5*) and Bell *et al.* (*19*)) is the population of scholars considered in the denominator which is the total population of scholars in the former and the mobile scholars in the latter. Since count of scholars in a year could be prone to under- or over-counting, using the count of mobile scholars entering a region or exiting would provide an alternative measure of exposure.

2.3 Net internal and international scholarly migration rates for an exemplar country

Here we present further results based on net scholarly migration rates at internal and international levels aggregated in sub-national regions for the 24 years of observation. We present trends for all states in the USA (Fig. S6) as one illustrative example to show how NMR rates change for these states over time. NMR rates presented further highlights the use cases of our database to identify sending and receiving sub-national regions.

2.4 More on Aggregated Net Migration Rate, and Migration Effectiveness Index measures

While the net migration rate provides a clear view of the number of scholars per 1,000 population of scholars who have entered or exited a region, it does not provide a sufficient basis for comparison between regions in different countries. Bell *et al.* (*19*) suggest using its aggregated version (ANMR) and Migration Effectiveness Index (MEI) that allows a better evaluation of how effective migration is in redistributing the population (based on the population at risk in the former case and among the mobile population in the latter case). Nevertheless, Bell *et al.* (*19*) highlight the shortcomings that might arise due to the number of regions in each country and the spatial dis-aggregation. We use GeoNames *first* administrative level for all countries to have a consistent view of these countries. However, it is still prone to shortcomings as some countries are divided into many sub-national regions. In addition, some countries (for instance the USA, the UK, Germany and Western, educated, industrialized, rich and democratic (WEIRD) countries in general) have a more developed science system and a higher number of institutions that allows more dynamic mobility of scholars.

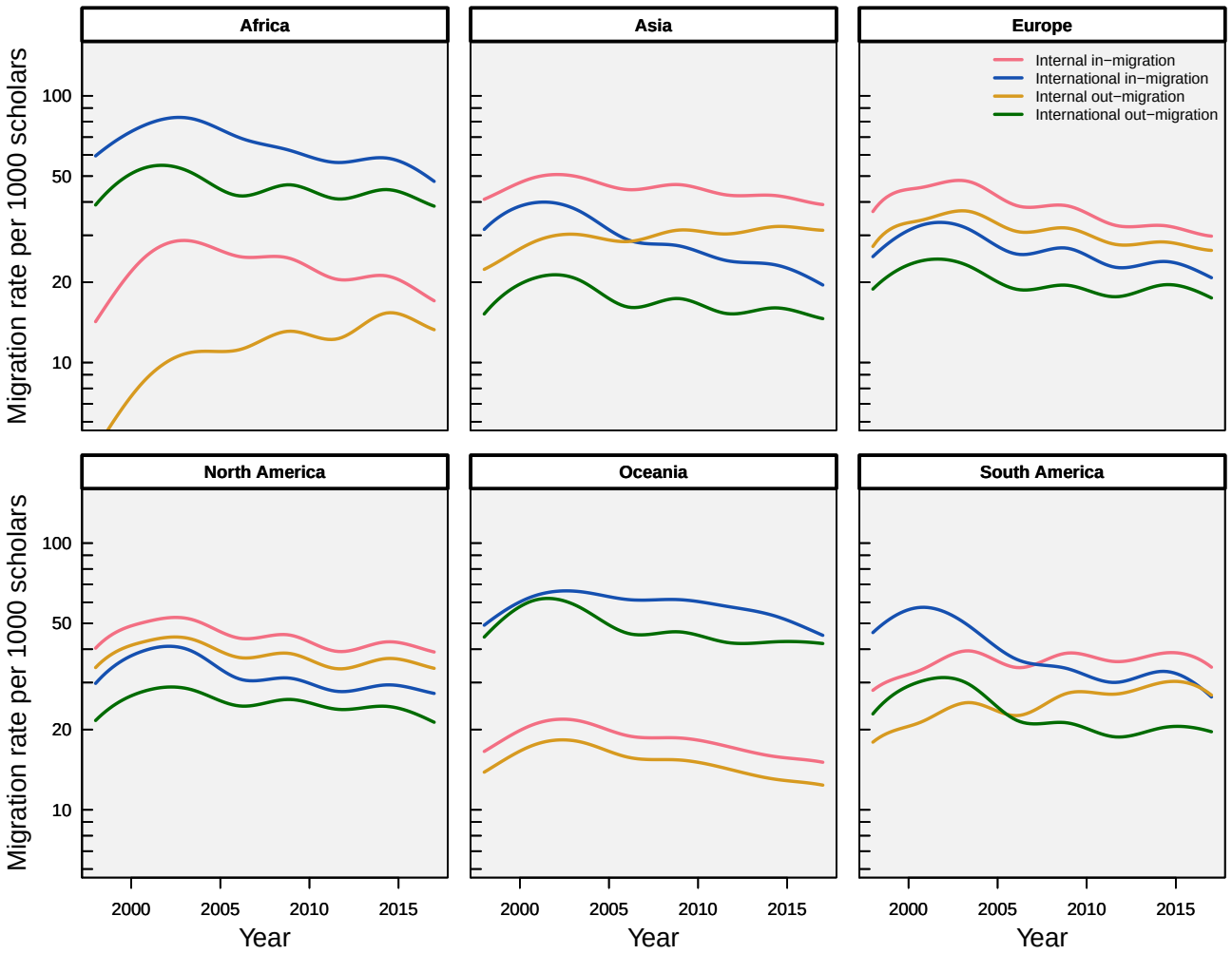


Fig. S1: **Predicted migration rates conditional on random effects for different types of migration, year, and continent, estimated via separate GAMMs indicated with line colors.** Each line represent an independent model visualized together to show temporal trends. While random effects of regions are included in models, however, they are excluded from the model predictions.

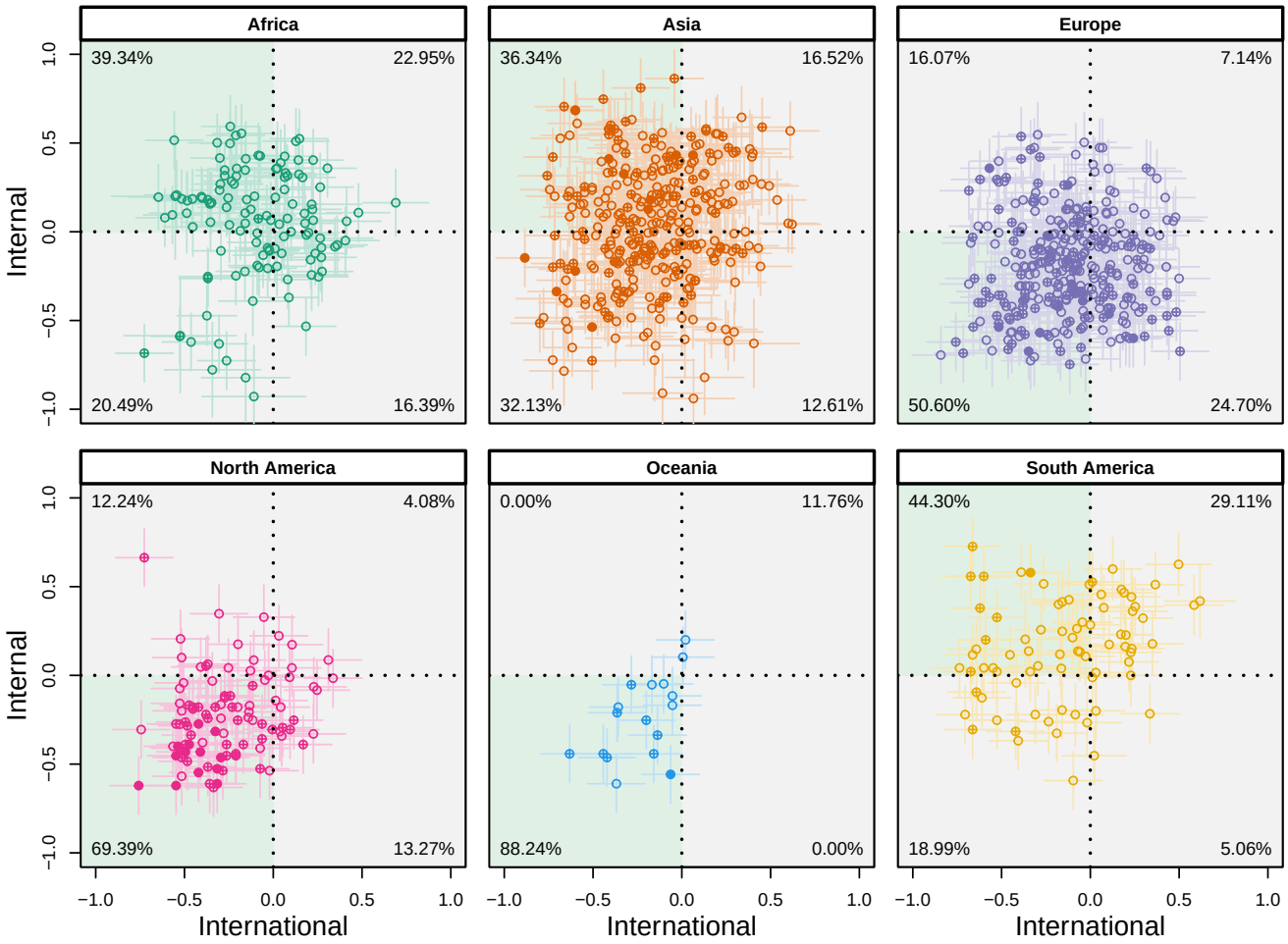


Fig. S2: **Kendall correlation coefficients for in-migration rate time trends for different regions (circles) split by continents (6 panels)**. Filled circles: population of scholars higher than 10,000; semi-filled circles: population of scholars between 1000 and 10,000; open circles population of scholars equal or lower than 1000. Percent values in each corner denote a fraction of cases per each quadrant. Upper left quadrant: international in-migration rate is negatively, but internal in-migration rate is positively correlated with year; Upper-right quadrant: both international and internal in-migration rates are positively correlated with year; Bottom-left quadrant: both international and internal in-migration rates are negatively correlated with year; Bottom-right quadrant: international in-migration rate is positively, but internal in-migration rate is negatively correlated with year.

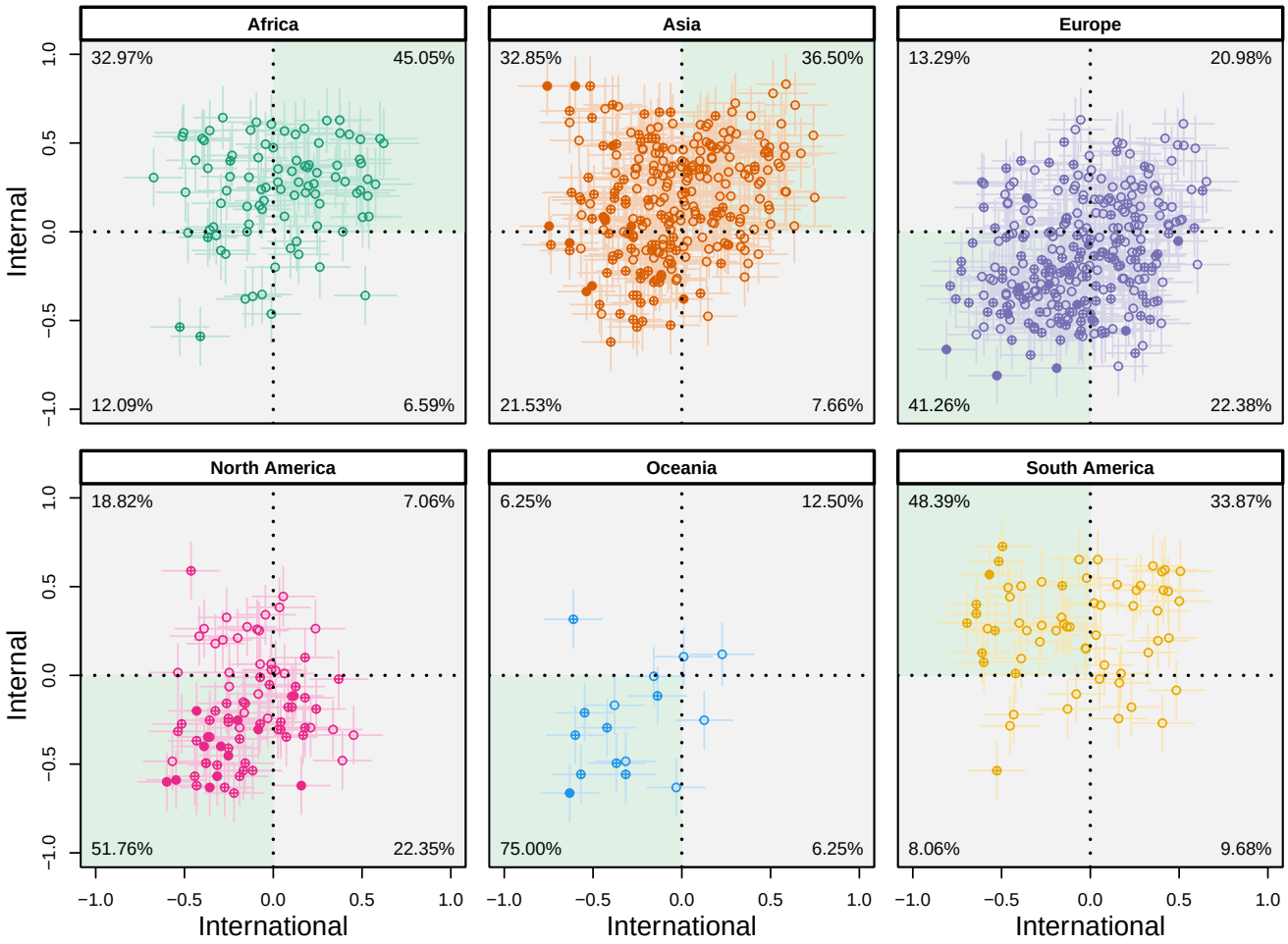


Fig. S3: Kendall correlation coefficients for out-migration rate time trends for different regions (circles) separated by continents (6 panels). The description of this figure is analogous to Figure 2

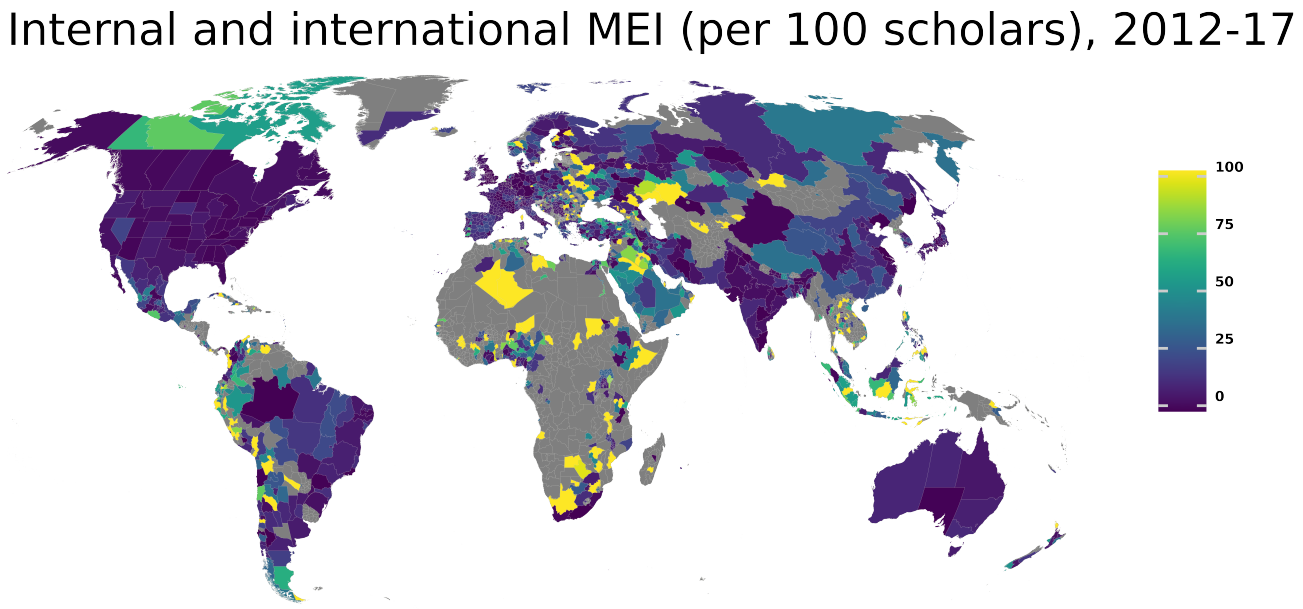


Fig. S4: **Combined internal and international Migration Effectiveness Index (MEI) per 100 scholars at the province level worldwide**. MEI ranges from 0 to 1 and indicates the overall rate of movements. Lighter and yellower colors show the regions with higher fluctuating rate of movement and darker and bluer colors show regions with more stable migration flows.

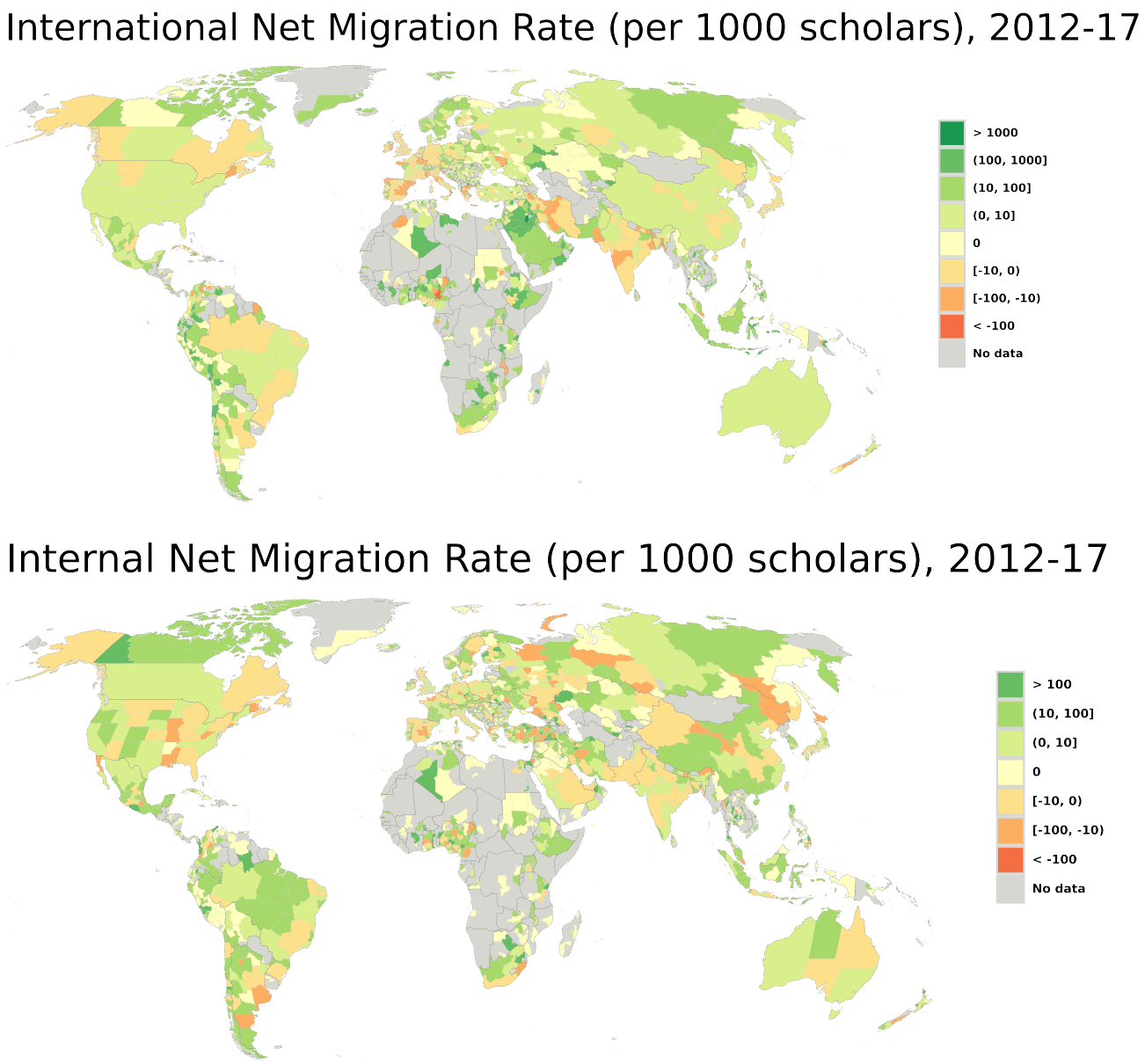


Fig. S5: **International (*A*) and internal (*B*) net migration rate per 1000 scholars at the province level**. Colors present the net migration rate of scholars sent (negative, red colors) or received (positive, green colors), and yellow shows a balanced flow. Color scales are kept similar in all maps using 8 classes to allow comparison while the upper and lower class includes above 1000 and below -100 net rates, respectively. Numbers printed on the legend are the net rate of scholars sent or received from/to international origins/destinations (A) or internal ones (B).

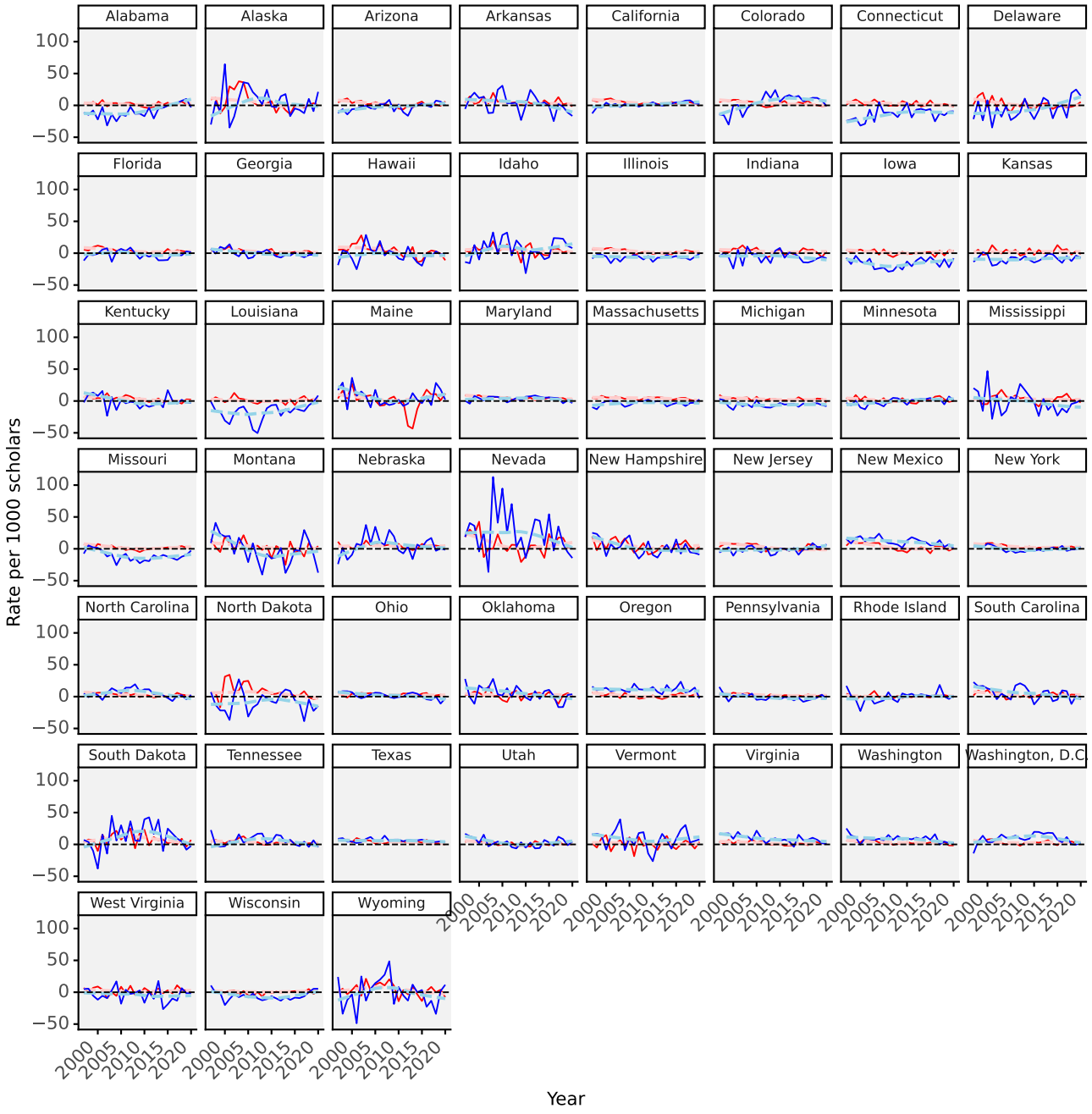


Fig. S6: Net internal (blue) and international (red) scholarly migration rates of US states (dashed lines are smooth trends based on mean). The first observation is the larger range of internal scholarly mobility versus international one (that causes the Y axis of the figure to have a longer range, i.e., -50 to 100). Some states are *sending* states in internal mobility (such as Alabama, Connecticut, Delaware, Iowa, Kansas, Louisiana, and Missouri), while some are *receiving* states (such as Colorado, Nevada, New Mexico, Oregon, South Dakota, Virginia, Washington, and Washington, D.C.). In international mobility, most regions are close to the zero line (which is mainly due to the higher range of internal mobility causing these trends to be less clear) except Alaska, North Dakota and South Dakota which are *receiving* states. Nevada has a high variation in both internal and international scholarly mobility. (Note that seven states out of 50 are missing from the figure as they did not have representation in our results).

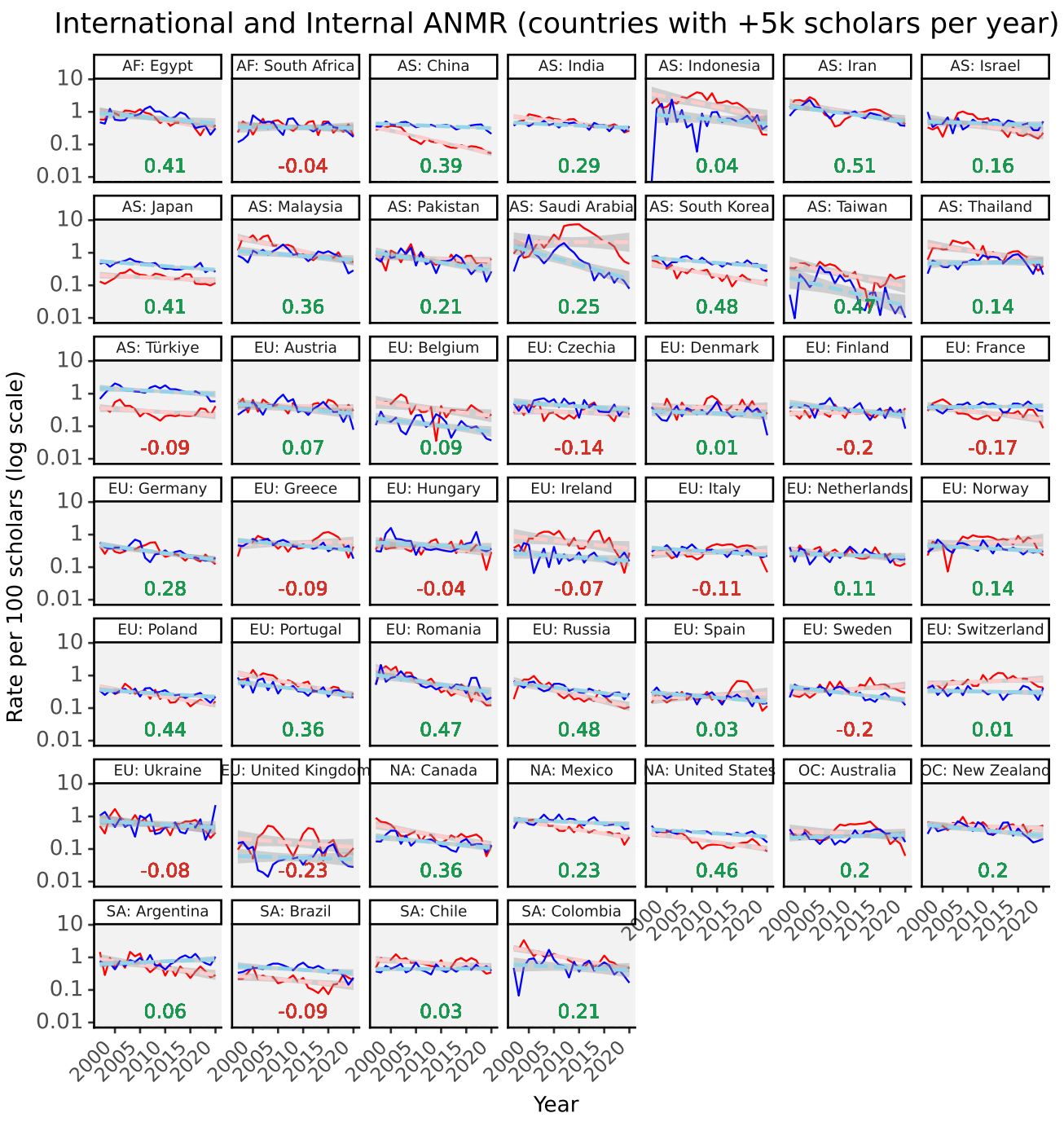


Fig. S7: **Countries with more than five thousand scholars per year and internal (blue, dashed line with lighter color for the smooth trend based on mean with confidence intervals in gray) and international (red) scholarly mobility and its comparative and aggregate view in terms of aggregate net migration rate (i.e., ANMR) per 100 scholars**. Numbers printed in bottom part of each panel is Kendall correlation of the internal and international ANMR over the 24 years of observation indicated in green (positive correlation) or red (negative correlation). This figure provides a nuanced *temporal* view of the continent (panel labels) and country differences in ANMR measure where some countries have a dominance of internal and some international scholarly migration that is stable over 24 years of observation, but in general most countries have declining or stable scholarly migration. 13 countries, most of which in the EU had negative correlations between two migrations.

References

1. German Competence Network for Bibliometrics, *Kompetenzzentrum Bibliometrie (KB)*, 2021, (2021; https://bibliometrie.info/).
2. A. Akbaritabar, *Quantitative Science Studies* **2**, 753–777, ISSN: 2641-3337, (2021; https://doi.org/10.1162/qss_a_00131) (2021).
3. E. Sanliturk, E. Zagheni, M. J. Dańko, T. Theile, A. Akbaritabar, *Proceedings of the National Academy of Sciences* **120**, e2217937120, (2023; https://www.pnas.org/doi/10.1073/pnas.2217937120) (Jan. 2023).
4. X. Zhao, A. Akbaritabar, R. Kashyap, E. Zagheni, *Proceedings of the National Academy of Sciences* **120**, e2214664120, (2023; https://www.pnas.org/doi/10.1073/pnas.2214664120) (Mar. 2023).
5. A. Miranda-González, S. Aref, T. Theile, E. Zagheni, en, *EPJ Data Science* **9**, Number: 1 Publisher: Springer Berlin Heidelberg, 34, ISSN: 2193-1127, (2021; https://epjds.epj.org/articles/epjdata/abs/2020/01/13688_2020_Article_252/13688_2020_Article_252.html) (2020).
6. A. Akbaritabar, T. Theile, E. Zagheni, "Global flows and rates of international migration of scholars", en, tech. rep. WP-2023-018, Edition: 0 (Max Planck Institute for Demographic Research, Rostock, Apr. 2023), WP–2023–018, (2023; https://www.demogr.mpg.de/en/publications_databases_6118/publications_1904/mpidr_working_papers/global_flows_and_rates_of_international_migration_of_scholars_7729).
7. B.-C. Björk, D. Solomon, en, *Journal of Informetrics* **7**, 914–923, ISSN: 1751-1577, (2022; https://www.sciencedirect.com/science/article/pii/S1751157713000734) (2013).
8. N. Robinson-Garcia *et al.*, *Journal of Informetrics* **13**, 50–63, ISSN: 1751-1577, (2021; https://www.sciencedirect.com/science/article/pii/S1751157718300865) (2019).
9. M. van Ham, en, *International Journal of Population Geography* **7**, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002 295–306, ISSN: 1099-1220, (2022; https://onlinelibrary.wiley.com/doi/abs/10.1002/ijpg.225) (2001).
10. R. Skeldon, en, *Population, Space and Place* **12**, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/psp.385, 15–30, ISSN: 1544-8452, (2021; https://onlinelibrary.wiley.com/doi/abs/10.1002/psp.385) (2006).
11. R. King, R. Skeldon, en, *Journal of Ethnic and Migration Studies* **36**, 1619–1646, ISSN: 1369-183X, 1469-9451, (2021; http://www.tandfonline.com/doi/abs/10.1080/1369183X.2010.489380) (2010).
12. E. Pedersen, D. Miller, G. Simpson, N. Ross, (https://doi.org/10.7717/peerj.6876) (2019).
13. R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing (Vienna, Austria, 2023), (https://www.R-project.org/).
14. S. N. Wood, *Generalized additive models: an introduction with R* (CRC press, 2017).
15. M. G. Kendall, *Biometrika* **30**, 81–93, ISSN: 00063444, (2023; http://www.jstor.org/stable/2332226) (1938).
16. J. B. Lewis, D. A. Linzer, *Political Analysis* **13**, 345–364, (https://EconPapers.repec.org/RePEc:cup:polals:v:13:y:2005:i:04:p:345-364_00) (2005).
17. P. Donner, C. Rimmert, N. J. van Eck, *Quantitative Science Studies* **1**, 150–170, (2020; https://doi.org/10.1162/qss_a_00013) (Dec. 2019).
18. J. Baas, M. Schotten, A. Plume, G. Côté, R. Karimi, *Quantitative Science Studies* **1**, 377–386, ISSN: 2641-3337, (2021; https://doi.org/10.1162/qss_a_00019) (2020).
19. M. Bell *et al.*, en, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **165**, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-985X.t01-1-00247, 435–464, ISSN: 1467-985X, (2021; https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-985X.t01-1-00247) (2002).