The Authors of Economics Journals Revisited: Evidence from a Large-Scale Replication of Hodgson & Rothman (1999)

Matthias Aistleitner[1], Jakob Kapeller[1,2]*, and Dominik Kronberger[1]

[1]Institute for Comprehensive Analysis of the Economy (ICAE), Johannes Kepler University Linz, Austria, [2]Institute for Socio-Economics, University of Duisburg-Essen, Germany
*Corresponding author. Email: jakob.kapeller@uni-due.de

Supplementary Information

1. Introduction

This file provides complementary information to the article "The Authors of Economics Journals Revisited: Evidence from a Large-Scale Replication of Hodgson & Rothman (1999)" published in the *Journal of Institutional Economics.* The purpose of this file is to provide information on specific methodological aspects (section 2) as well as to present further and more detailed graphical outputs of the data used in our study (section 3).

2. Data and Methods: selection of journals and data processing

2.1 Defining a set of top journals

To define a definite set of thirty journals in the course of our study we proceeded in the following way: first used the Journal Citation Reports (JCR) to identify the thirty most influential journals in economics considering 22 consecutive years (1997-2018, JCR category: economics). To account for the relative long-term position of outlets, we inspected the JCR rankings published in the years under study, counted how often journals were ranked in the top 30 and included all journals above a certain threshold. As we aimed to select a number as close as possible to thirty top journals, we treated the threshold value as a floating variable to eventually arrive at a selection of 30 journals (which led us to a minimal threshold of at least nine appearances in the top 30)[1]. While our selection criterium is quite different from that of H&R it should be noted that the overlap between both journal sets is substantial (21 out of 30 journals) which reflects a strong persistence of dominant outlets within the discipline.

Table S1 provides an overview of the journals in our sample which includes top generalist economic journals as well as a wide range of economic field journals (e.g. finance, accounting, health and environmental economics)

[1] Since journals #29-31 appeared with the same frequency in the top 30, we excluded the journal with the lowest average ranking.

*Table S1. Dataset: selected journals and the overall number of analyzed papers. Footnotes denote the overlap with the top 30 in journal sets analyzed by *Hodgson and Rothman (1999), [1]Kalaitzidakis et al. (2003), and [2]Kalaitzidakis et al. (2011)*

Rank	Journal	# papers
1	Journal of Economic Literature*,1,2	634
2	Quarterly Journal of Economics*,1,2	1,254
3	Journal of Economic Perspectives*,1,2	1,471
4	Journal of Political Economy*,1,2	1,335
5	Journal of Financial Economics*,1,2	2,232
6	Econometrica*,1,2	1,759
7	Economic Geography*	543
8	American Economic Review*,1,2	5,681
9	Review of Economic Studies*,1,2	1,313
10	Review of Economics and Statistics1,2	2,250
11	Journal of Accounting & Economics*	909
12	Ecological Economics*	4,448
13	Journal of Economic Growth2	311
14	Brookings Papers on Economic Activity*,2	1,012
15	Journal of Health Economics*	1,795
16	Economic Journal*,1,2	2,539
17	Journal of Economic Geography	653
18	Health Economics	2,143
19	Journal of Environmental Economics and Management*,1	1,519
20	Journal of International Economics1,2	1,820
21	Energy Economics	3,115
22	Journal of Labor Economics*,1,2	889
23	Journal of Law & Economics*	811
24	Journal of Finance	2,264
25	Journal of Human Resources*,1,2	1,043
26	Journal of Monetary Economics*,1,2	1,961
27	Economic Policy	1,018
28	Journal of Law Economics & Organization*	723
29	Review of Environmental Economics and Policy*	218
30	Review of Financial Studies	1,806
	Total	**49,469**

2.2 Disambiguation of textual data

The analysis of the bibliometric data in a way suitable for our purpose comes with a series of practical challenges mostly related to the processing and disambiguation of text-based data. The two most important issues in the context of our data – which is based on EconLit – relate to the disambiguation of names of academic institutions and journal authors. The former incorporates issues such as distinguishing between *and* as part of a single affiliation (e.g. "London School of Economics *and* Political Science") and as a conjunction (e.g. "Harvard *and* NBER"), which posed a specific challenge. In this context, we set the interpretation of *and* as a conjunction as a default and compiled a go-list with institutions including the phrase *and* by consulting the EDIRC database

provided by RePEc[2] to avoid errors when processing author affiliations. Another related problem are commas that can be either part of an affiliation name or are used to list multiple affiliations of a single author. Since in most cases, authors do not list more than two affiliations we assume that commas are part of an institutions name (e.g. "Environmental Policy Centre, Finnish Environment Institute, Helsinki"). Inspections of repeated random samples drawn from the data show that this strategy results in an error rate of around 2% of all compiled affiliations.

Furthermore, since our analysis is based on plain textual data, name variation and disambiguation remain fundamental problems that could only be partially addressed in the course of our study. To confront this issue, we manually checked and harmonized different spellings of the names of the top 1,000 authors (in terms of authorships) and affiliations respectively[3]. Finally, we also eliminated entity ambiguity in author for which we manually collected information on their PhD-granting institution.

2.3 Collection of country information for the analysis of the geographical distribution of author affiliations

Finally, we collected geographical data to complement our full dataset by a spatial dimension via a two-step procedure. First, we match country information available from the random sample of PhD affiliations with the affiliations contained in the overall dataset. Second, we manually gathered country information from all remaining affiliations that appear at least three times in the dataset (which results in a coverage rate of 89% of all author-affiliation pairs).

3. Results

In this section we use descriptive statistics to document findings based on our data that are mentioned throughout the text. The graphical outputs here serve not only to document the empirical plausibility of our verbal claims but also could prove useful to the interested reader as they provide a more exhaustive and precise account of some of our main results and, hence, substantiate and complement the findings reported in our paper.

[2] See also https://edirc.repec.org/.

[3] It should be noted that there are some (rare) cases where two authors have an identical name which is difficult to address in the full dataset. In the case of the PhD sample, we excluded such rare cases.

3.1. Geographical origin of authors

This section graphically documents the geographical diversification of authors measured by institutional origin (affiliations in Figure S1, PhD-granting institutions in Figure S2).

Figure S1. Geographical distribution of affiliations over time.

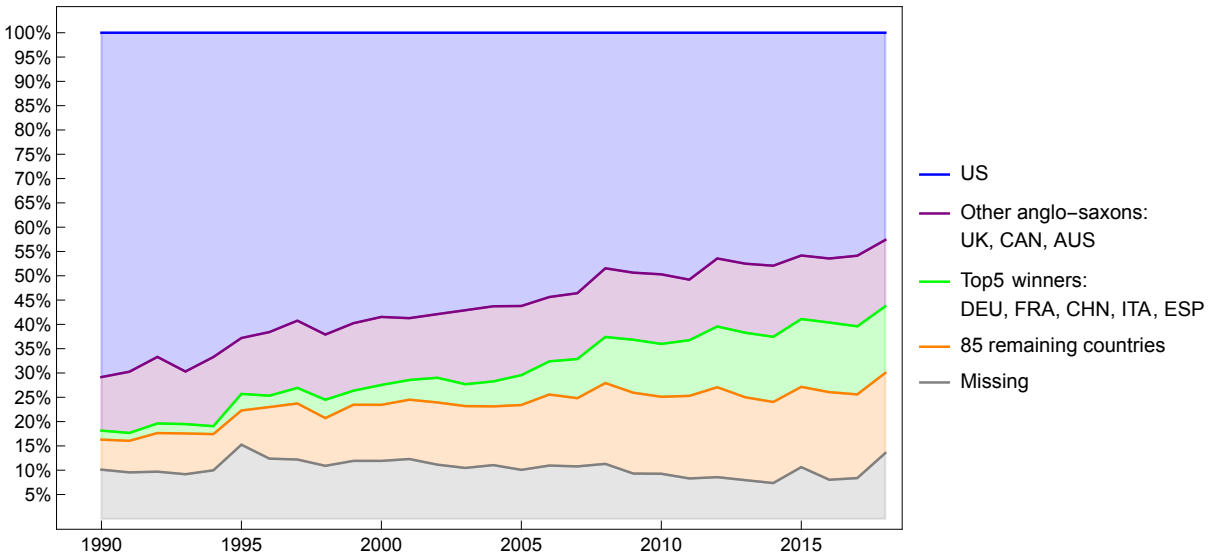
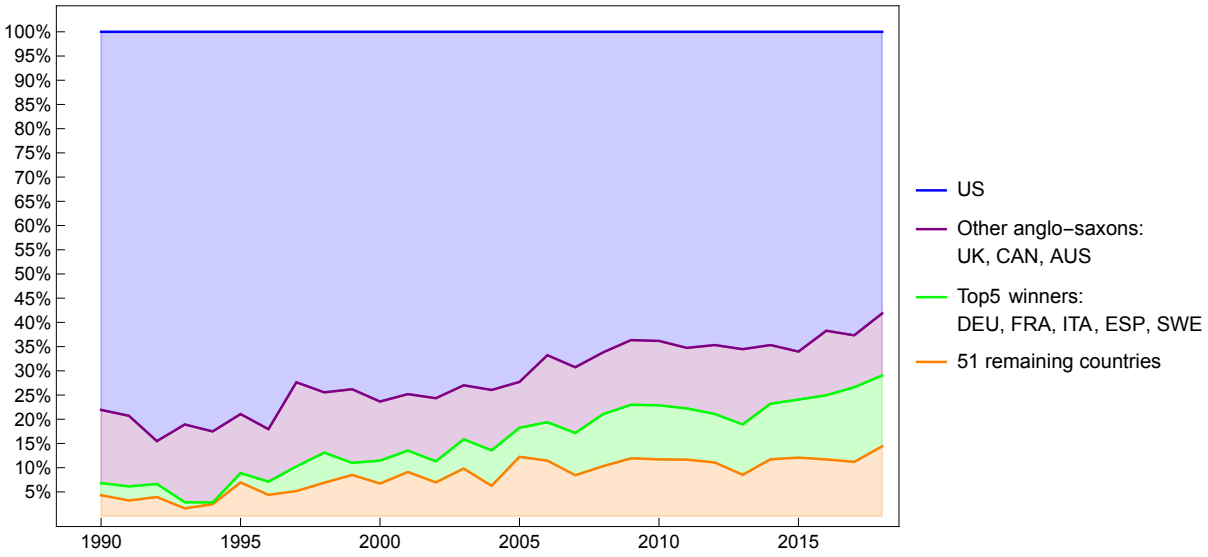


Figure S2. Geographical distribution of author’s PhD-granting institutions over time.



3.2. Complementary information on the comparison with H&R (1999)

Here we provide two plots complementing our comparison with H&R (1999) as presented in the main paper. For the most visible institutions we show ranks #16-30. (Figure S3), for the most prominent PhD-granting institutions ranks #1-20 (Figure S4), within the top 30 journals as defined by H&R (left bars) and in the replication data

(middle and right bars). As emphasized in the correlation between the original results and the results obtained from the replication data decreases for the bottom half of the top 30 institutions as identified by H&R, which is not all too surprising given the differences in terms of sample composition.

Figure S3. Most important affiliations in top 30 journals: H&R vs. replication. Ranks #16-30. The affiliations are ordered according to their relative share in the replication study.

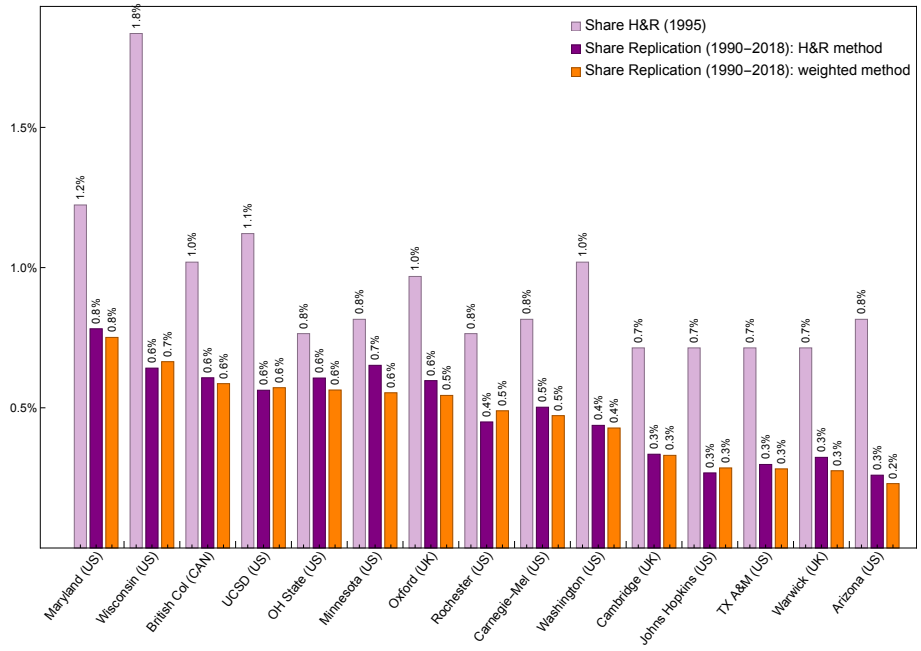
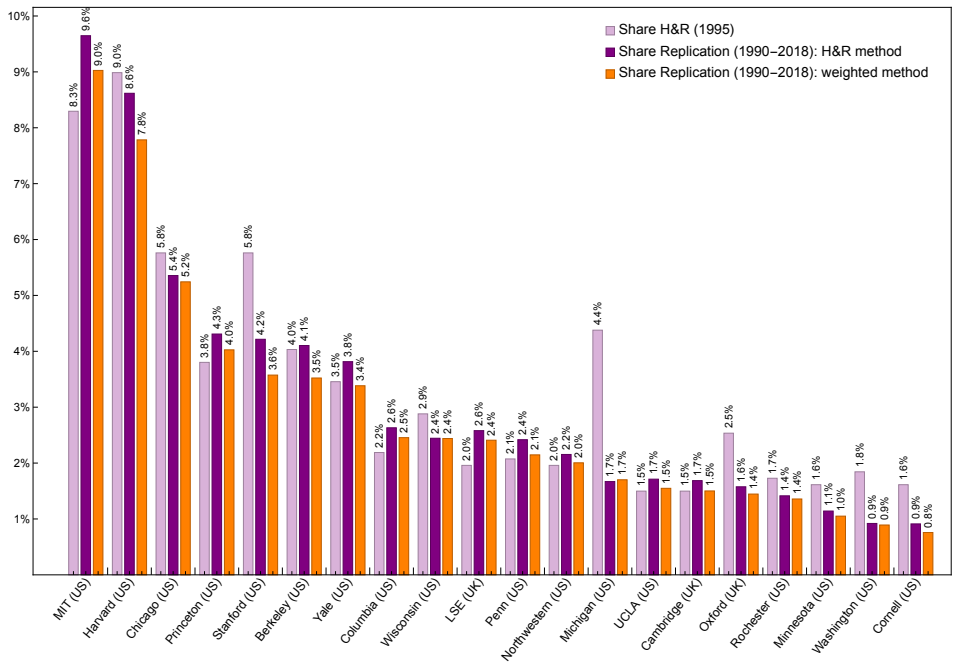


Figure S4. Most important PhD-granting institutions in top 30 journals: H&R vs. replication. Ranks #1-20. The affiliations are ordered according to their relative share in the replication study.



3.3. Evaluating concentration in author affiliations on the level of journals

In this section we document institutional concentration on the level of journals for all thirty journals under study by means of (a) barcharts of most important institutions within a given journal (Figures S5a and S5b) (b) Lorenz-curves measuring institutional diversity within journals (Figure S6).

Figure S5a. Most prominent/important author affiliations among the top 30 journals: journal-specific results.

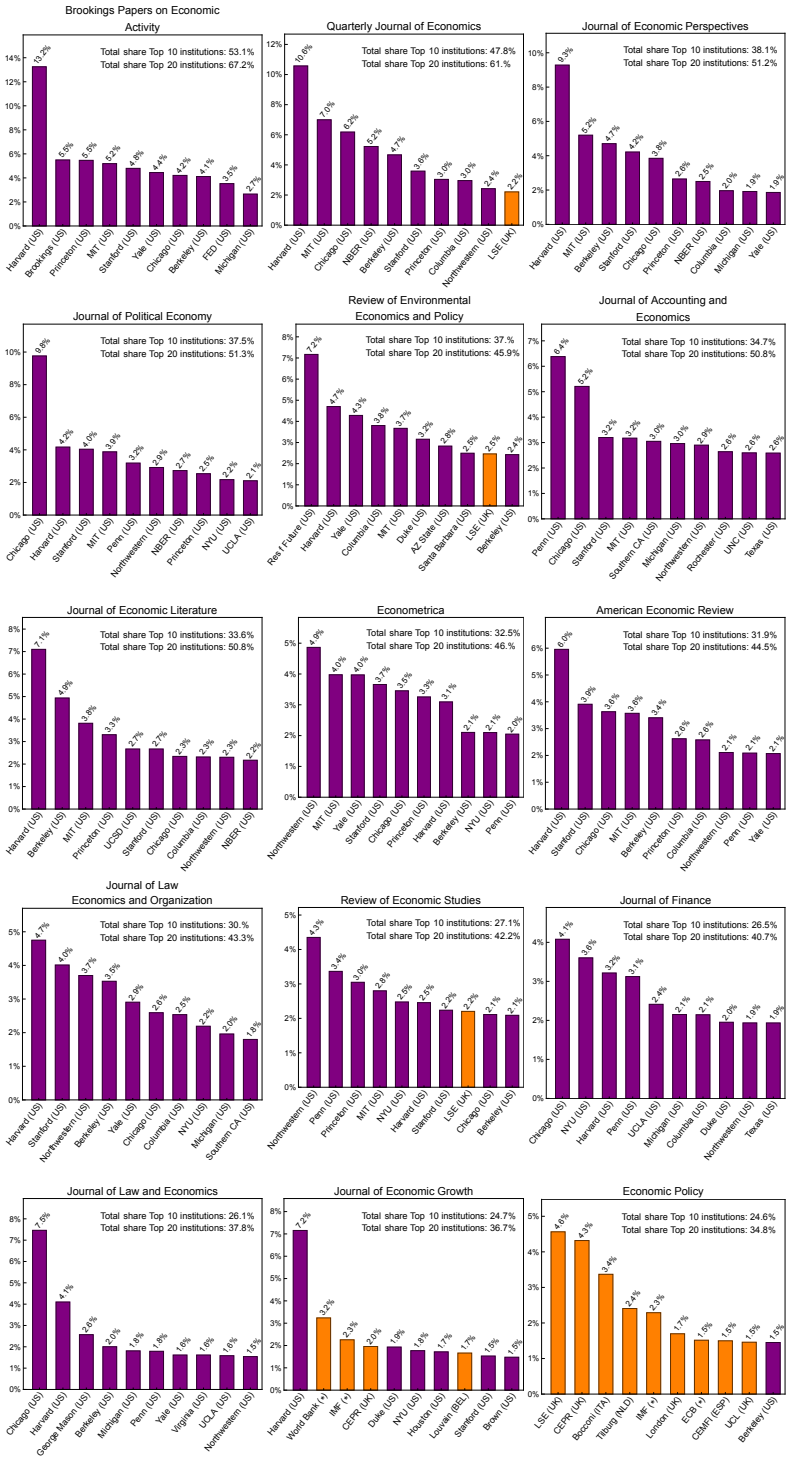
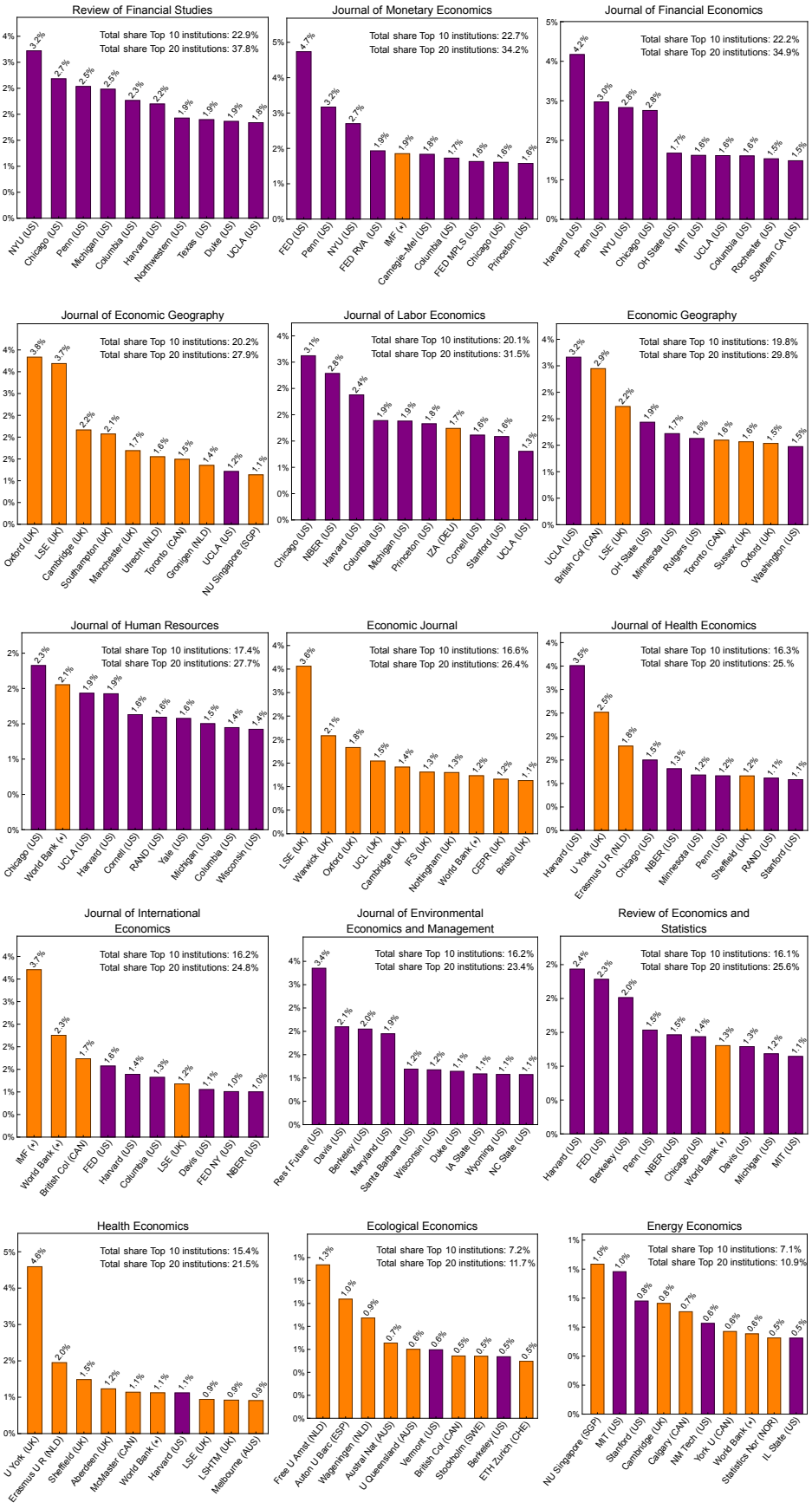


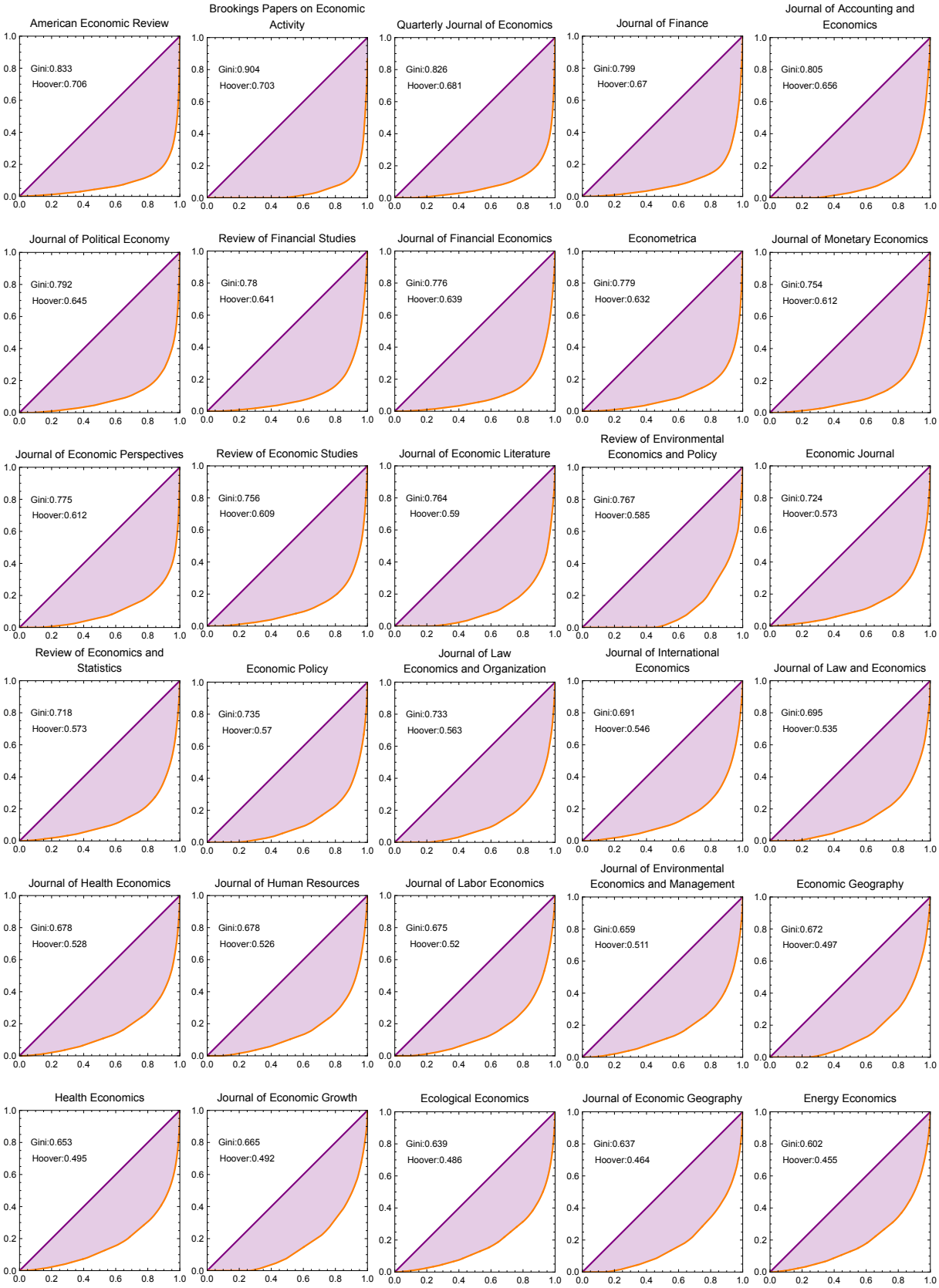
Figure S5b. Most prominent/important author affiliations among the top 30 journals: journal-specific results.



In what follows, we present Lorenz-curves, Gini-coefficients and Hoover-Indices for all journals under study, where the ordering of journals is governed by the Hoover-Index. A key difference between this approach to the one mainly used in our paper – bar-charts emphasizing the role of top institutions – is that the former is more sensible with regard to how the underlying population is defined. In other words, the question is about how to treat zeros in our data that appear when some institution is not at all represented in a given journal. As the bar-charts focus on top institutions in absolute terms (‘top 10’), they are not much affected by the question how many institutions are represented in a given journal. However, the Gini as well as the Hoover Index – the latter is defined as the share of journal space that would need to be redistributed to ensure equal representation of all institutions – care for relative values, which makes it necessary to define some relevant population. To capture the underlying intuition, consider the hypothetical case, in which only ten institutions publish in a journal and they do so in equal shares. As a result, a bar-chart based assessment would point towards extreme concentration (100% of all articles are related to the top 10 institutions), while both, Hoover as well as Gini, indicator would approach their lower bound (in relative terms, there is perfect equality between all institutions able to publish in this journal). To compensate for this fact and to make the resulting rankings better comparable across journals we defined the base population by taking the top 2% of all institutions in our dataset. We then complemented the list of institutions publishing in some journal by comparing it with our base-population to account for the degree of ‘total exclusion’ of institutions from some journals.

Hence, journals shown at the top-left of Figure S6 are more concentrated than those shown at the bottom right. Figure S6 thereby further confirms our initial impression – that highly visible, generalist journals tend to be more concentrated as all top five journals as well as the *Journal of Economic Perspectives* reside in the upper half of the resulting ranking. However, we also observe that the field of finance differs from the other research fields as finance-journals also exhibit a higher degree of concentration according to this measure. Inversely, the bottom half of the resulting ranking is mainly populated by highly regarded field journals.

Figure S6. Lorenz-curves, Gini-coefficients and Hoover-Indices for all journals under study



The value of the Gini-Index is above 0.8 for the most unequal journals, while the lowest values are slightly above 0.6. While such comparisons should be taken with a grain of salt, the upper range of these values is equivalent to Gini-coefficients found in analyses of wealth distributions, which can be considered as one of the most strongly skewed distribution to be found in economic data. Also, the lower range of the observed values is somewhere in-between of the values typically obtained from wealth-data and data on income inequality in developing countries. As the latter are more unequal than the typical developed country, all our estimates can be considered to signify far greater inequality than that expected of the income distribution of a developed country (which is already coined by a strikingly high degree of absolute and relative differences in income).

3.4. Evaluating concentration of PhD-granting institutions on the level of journals

Similar to Figures S5a and S5b, which showed detailed accounts for institutional concentration within journals as measured by author-affiliations, Figures S7a and S7b reproduce this exercise for examining the relative impact of different PhD-granting institutions across journals.

Figure S7a. Most prominent/important PhD-granting institutions among the top 30 journals: journal-specific results

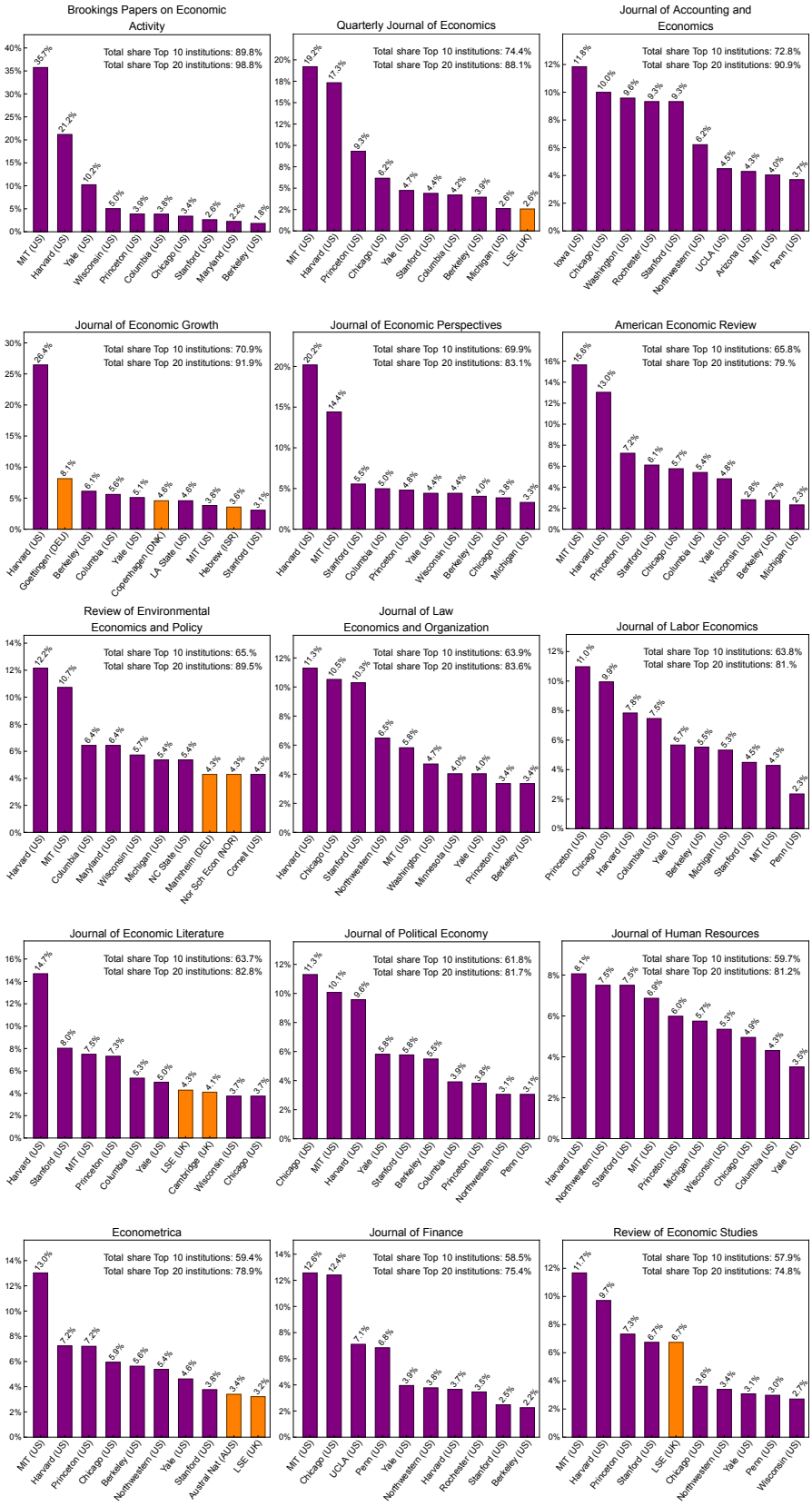
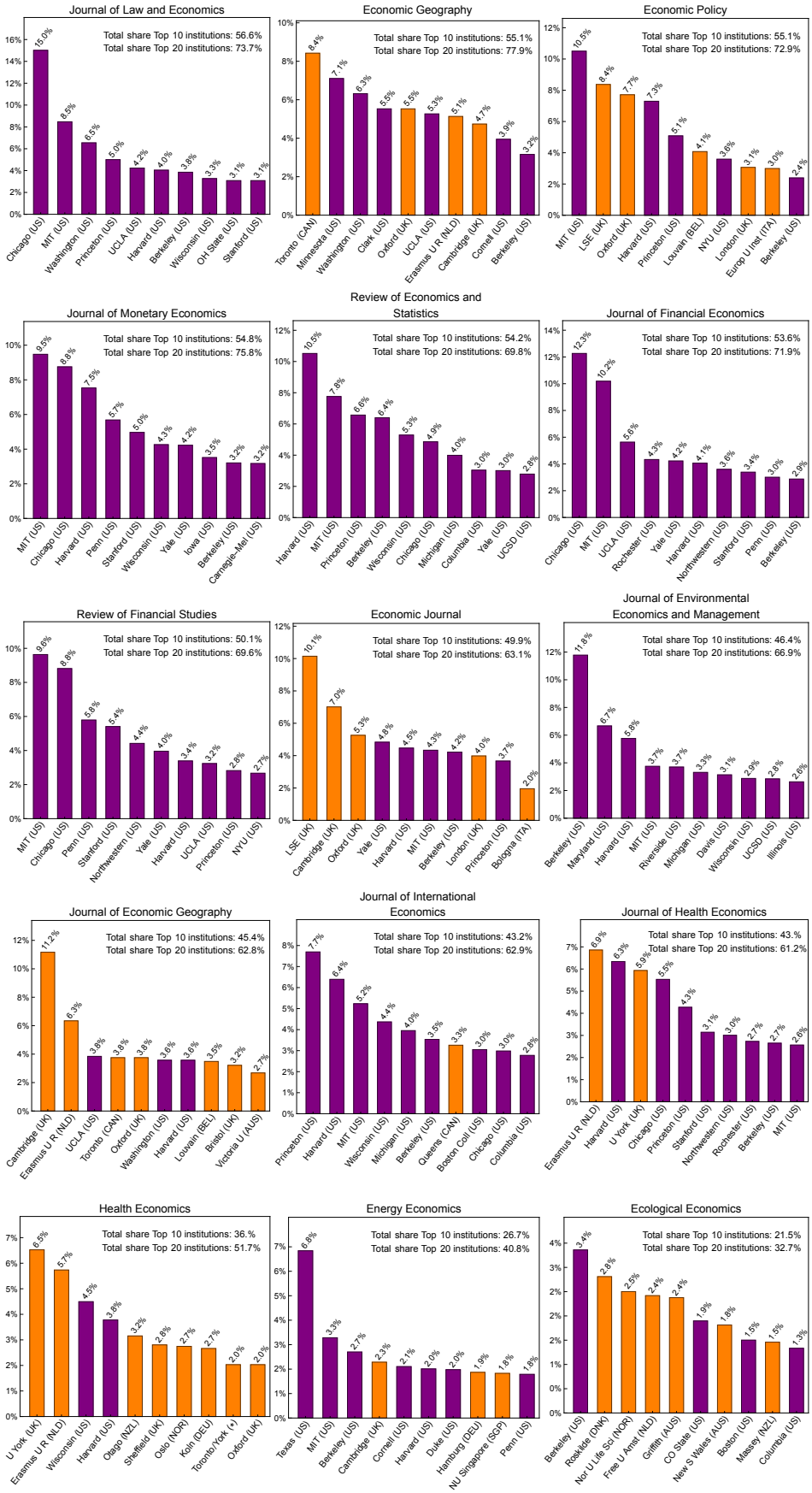


Figure S7b. Most prominent/important PhD-granting institutions among the top 30 journals: journal-specific results



3.5. A temporal perspective: evaluating institutional concentration over time

In this final part and complementary to Figures S1 and S2 (see section 3.1 above), we further explore the temporal properties of our dataset, to substantiate our claim that the phenomena described so far are, with some exceptions, indeed stable over time. Presenting the underlying information in a concise and yet accessible way, benefits from some additional aggregation on the side of journals. In our case we choose to plot the cumulative share based on three sets of top institutions (top 3/top 5/top 10) as given by a simple ranking of institutional visibility (also utilized in Figures 5 and 6 in the main paper). Basically, we ask whether the share of this homogenous group of top institutions represents a stable hierarchical position has – that is, a stable share in articles in top-journals – over time.

Figure S8. Comparing concentration of top institutions (author affiliations) in top 5 and top 30 journals over time

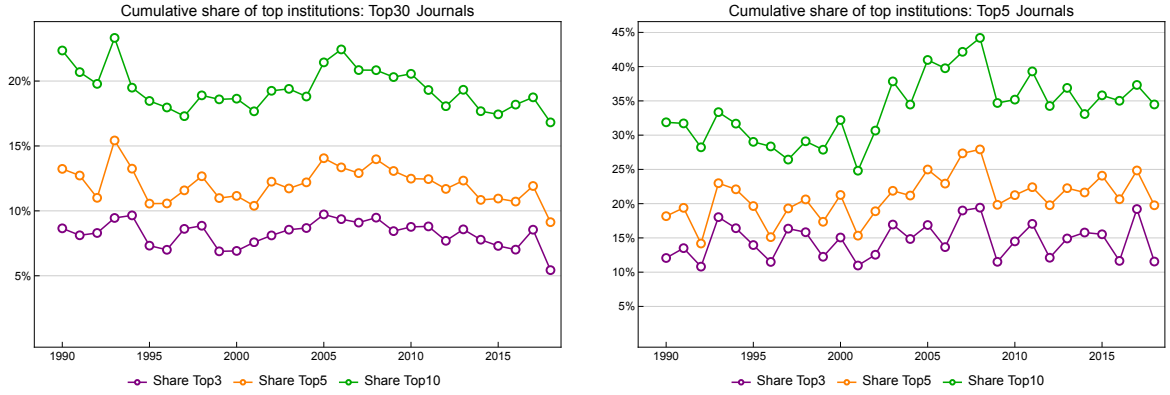
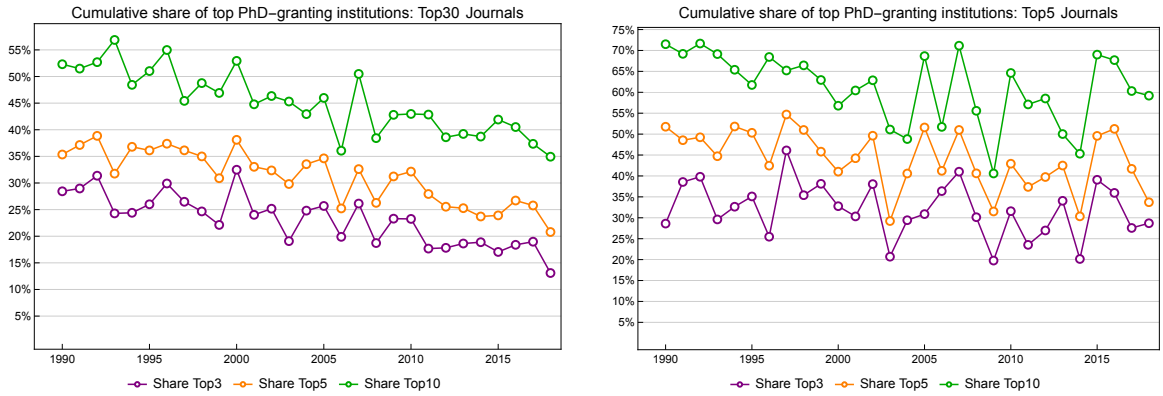


Figure S8 provides a first answer to the question posed above: while in the top 30 journals the share of top institutions appears to be slightly decreasing in the past 15 years or so, the share of contributions emerging from the most dominant institutions in the top 5 journals remains quite stable for the past 15 years and has even seen a slight increase before that.

The same assertion holds, by and large, for the time-series view on PhD-granting institutions (see Figure S9). The one main difference is that we do indeed observe a slight downward trend of the impact of graduates from major PhD-granting institutions over time (although this development is starting from a very high level). However, if we focus solely on the top 5 journals instead this trend becomes considerably weaker – especially for the very top institutions.

Figure S9. Comparing concentration of PhD-granting institutions in top 5 and top 30 journals over time



4. References

Hodgson, Geoffrey M., and Harry Rothman. 1999. “The Editors and Authors of Economics Journals: A Case of Institutional Oligopoly?” *The Economic Journal* 109 (453): 165–86. https://doi.org/10.1111/1468-0297.00407.

Kalaitzidakis, Pantelis, Theofanis P. Mamuneas, and Thanasis Stengos. 2003. “Rankings of Academic Journals and Institutions in Economics.” *Journal of the European Economic Association* 1 (6): 1346–66. https://doi.org/10.1162/154247603322752566.

———. 2011. “An Updated Ranking of Academic Journals in Economics.” *Canadian Journal of Economics/Revue Canadienne d’économique* 44 (4): 1525–38. https://doi.org/10.1111/j.1540-5982.2011.01683.x.