The Evolving Economics Research Collaboration Patterns

Willy Chen
Michigan State

Xiao Qiao
City University of Hong Kong

Hanzhe Zhang
Michigan State

American Economic Association
2025

Motivation

- How academic research is conducted has massively changed in past decades
- As information technology advances, research collaboration has become simpler than ever
- Across fields, there are more multi-authored papers
  - Among others, Wuchty et al. (2007, Science) document increased collaboration in natural and social sciences
- Why do we collaborate more?
  - Jones (2021): Multi-authored papers are more likely to have high impact

Research Goals

- Document stylized facts in economics research collaboration
  - Improve breadth: Expand the existing analysis to more than three times the covered time period, starting in 1886
  - Go deeper: Current findings focus on how many people work together, while we look for more granular characteristics
- Test hypotheses about the drivers of observed trends
  - Why do we observe increased collaboration?
  - Greater returns to collaboration? Decreased costs?

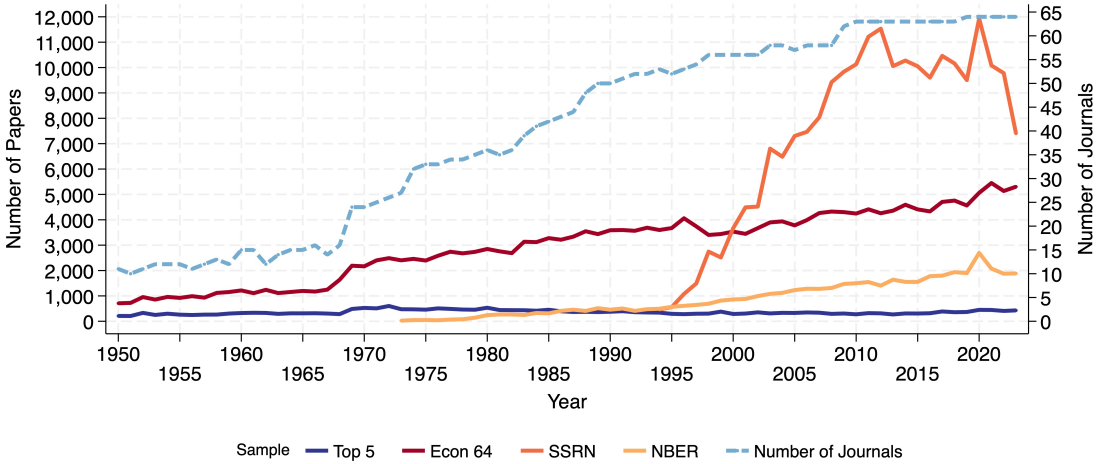
Our Findings

- The intensity of collaboration has increased over the past few decades
  - Fraction of multi-authored papers went up
  - The average number of authors on a paper went up
- Inter-institutional collaboration first decreased, then increased over time
- No change in how junior and senior economists work together
- A random utility framework to understand individual choice
- Regression evidence suggests that increasing returns may be a driver to increased collaboration
- Covid-19 as an exogenous shock to paper production
  - Some factors make collaboration more difficult: No conferences or research visits
  - Others make it easier to collaborate: Zoom, remote work
  - More single-authored and many-authored papers relative to historical trend

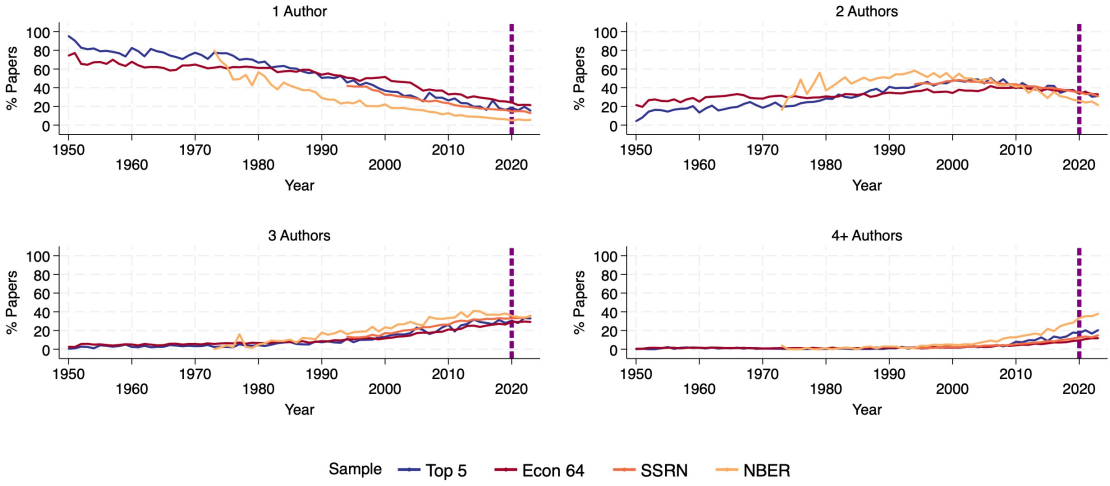
Data

- Four sets of papers
  - SSRN (1994 - 2023)
  - NBER working papers (1973 - 2023)
  - Top five journals (1886 - 2023)
  - 64 economics journals (1886 - 2023)
    - Includes top five, top field, etc.
- Author affiliation over time - hard!
  - Repositories only retain the latest affiliation
  - Difficult to study inter-institutional collaboration
  - We construct a time series of affiliations for each author using their publications
  - Spot check with academic CVs
- Main analysis centers around paper-year observations
  - Title, publication year, journal/repository, author names, affiliation, etc.

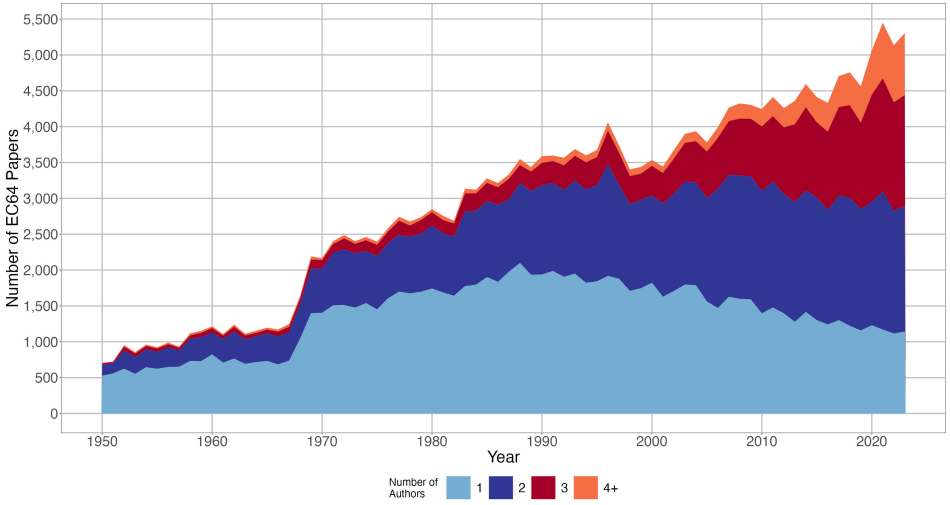
Growth of the Economics Field



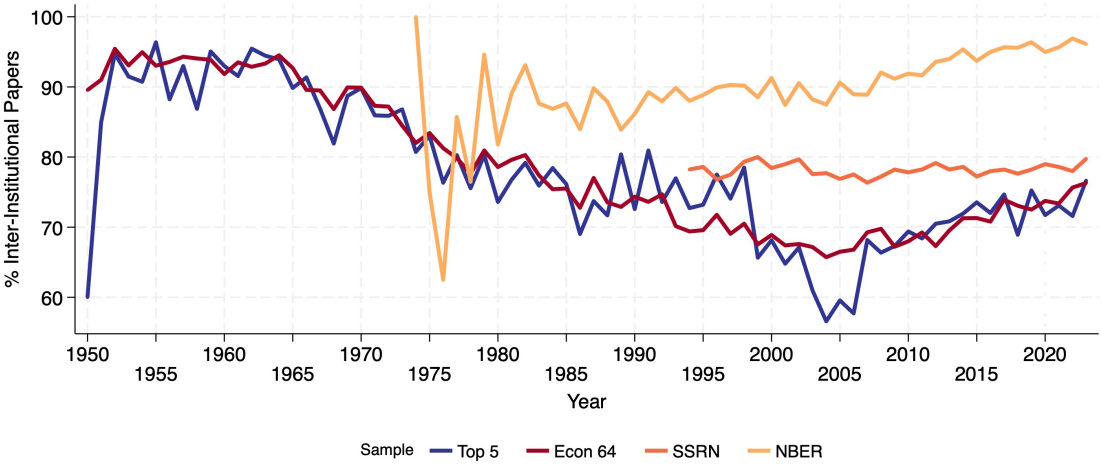
More Multi-Authored Papers Overall



Number of Authors for Econ64 Papers



Trends in Inter-Institutional Collaboration



A Random Utility Framework

Consider a simple random utility model (McFadden, 1974):

- r_1, r_m: The average return to a solo-authored paper and a multi-authored paper, the impact of a paper (citations as empirical proxy)
- c_1, c_m: The average cost per author of producing a paper of each type, includes all factors to the production of a paper such as time, opportunity, and financial costs
- $\varepsilon_{1,i}$, $\varepsilon_{m,i}$: Researcher i's idiosyncratic preference for each type of paper

The economist i would prefer to collaborate with someone instead of working alone if:

$$r_m - c_m + \varepsilon_{m,i} > r_1 - c_1 + \varepsilon_{1,i}.$$

Trends in Returns to Collaboration

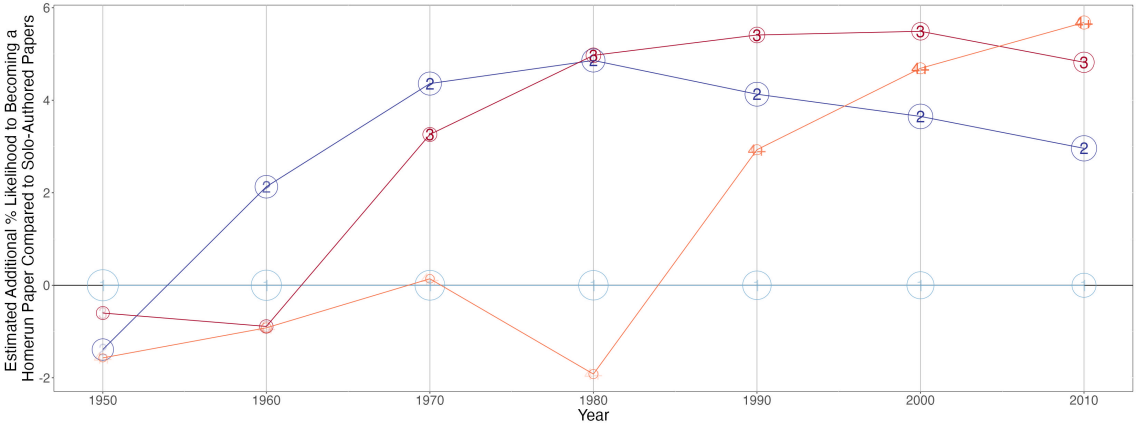
- *Homerun* (HR_{nit}): 1 if a paper i with n authors is in the top 10% of 5-year citations among papers published in the same year t

We estimate the following equation:

$$100 \cdot HR_{nit} = \alpha + \sum_{n} \beta_n + \gamma X_{it} + \kappa_t + \phi_j + e_{nit}$$

β_n, κ_t, and ϕ_j are n-author, year, and journal fixed effects, X_{it} include paper-level characteristics such as $\mathbb{1}\{\text{Has Senior Author}\}$, $\mathbb{1}\{\text{US Affiliated}\}$, and $\mathbb{1}\{\text{International Collab}\}$.

Increasing Returns to Collaboration



Note: Estimating equation is equation (6) with year and journal FEs for each decennial. Each tick on the x-axis represents the 10-year period starting that year.

Do Returns Drive Prevalence?

- $HRShr_{nt} = \sum_i HR_{nit} / \sum_i P_{nit}$: homerun papers with n authors as a fraction of all n-authored papers
- $RNS_{nt} = \sum_i P_{nit} / \sum_i P_{1,it}$: ratio of n-authored papers to solo-authored papers

We estimate the following system of equations under the seemingly unrelated regressions framework as proposed by Tomz et al. (2002):

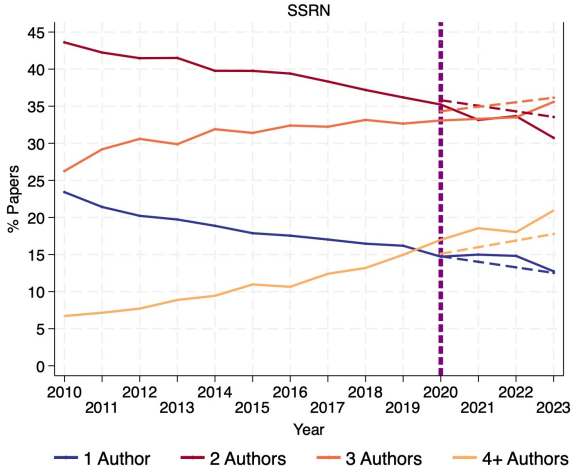
$$\Delta ln(RNS_{nt}) = \alpha_n + \gamma_n t + \beta_n ln(RNS_{n,t-5}) + \sigma_n ln\left(\frac{HRShr_{n,t-5}}{HRShr_{1,t-5}}\right) + \varepsilon_{n,t}$$

where β_n estimates how much economists converge to trend and σ_n estimates how much economists collaborate more due to increased returns.

Table 1: Regressions of Prevalence on Returns, 1950-2018

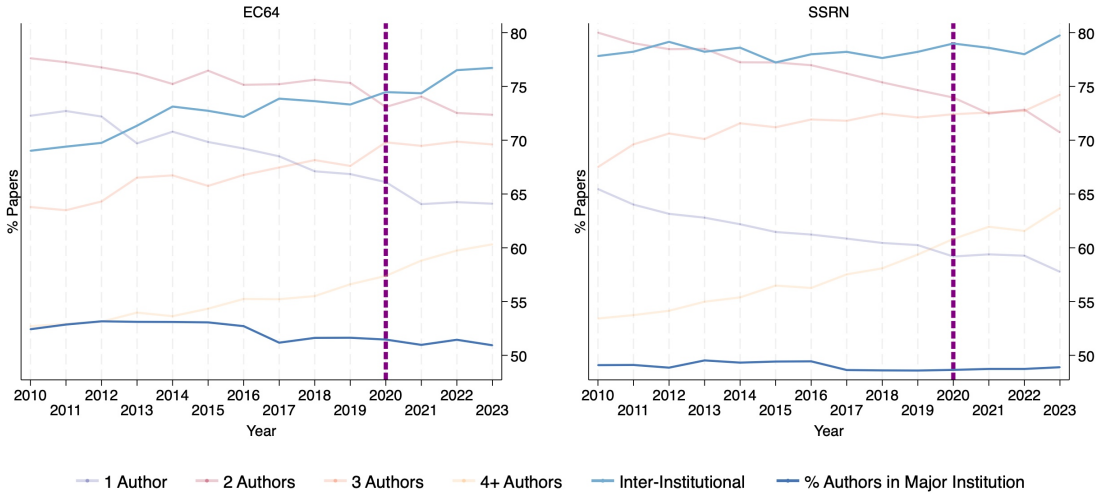
		$\Delta ln(RNS_{2t})$		
σ_2	-0.0068	-0.0293	-0.0615	-0.0748
	(0.0258)	(0.0364)	(0.0485)	(0.0525)
β_2		0.0416		-0.0780
		(0.0416)		(0.0809)
Year			0.0017	0.0034
			(0.0010)	(0.0021)
		$\Delta ln(RNS_{3t})$		
σ_3	0.0043	-0.0353	-0.0634*	-0.0638*
	(0.0244)	(0.0341)	(0.0382)	(0.0387)
β_3		0.0499		-0.0583
		(0.0310)		(0.0627)
Year			0.0033**	0.0056*
			(0.0014)	(0.0029)
		$\Delta ln(RNS_{4t})$		
σ_4	0.0690**	0.0744*	0.0666*	0.0712*
	(0.0347)	(0.0409)	(0.0387)	(0.0410)
β_4		0.0077		-0.0517
		(0.0508)		(0.0715)
Year			0.0011	0.0024
			(0.0016)	(0.0023)

How Did COVID Affect Collaboration?



Dashed lines are linear predictions.

How Did COVID Affect Collaboration?



Summary

- Expand and extend existing evidence on academic collaboration
- Multi-authored papers may be driven by increasing returns to collaboration
- Covid-19 serves as an exogenous shock that primarily shifts the costs associated with paper production

Looking ahead:

- Estimate a structural model of paper production
- Incorporate the draft-to-publication delay
  - Hadavand et al. (2024) note this delay is not related to greater academic attention
- More work needed to further understand drivers of successful collaboration

THANK YOU!

Appendix: Constructing Author-Affiliation-Year Record

Consider the case of an author who was ever affiliated with institutions A, B, and C:

For each author-affiliation pair, we record the first and last year that combination appears. We call these *maxyear* and *minyear*.

Affiliation	y2005	y2006	y2007	y2008	y2009	y2010	minyear	maxyear
A	1	1	.	.	.	.	2005	2006
B	.	1	1	.	.	1	2006	2010
C	.	1	.	.	1	1	2006	2010

Appendix: Constructing Author-Affiliation-Year Record

Step 2: For every year that is between *maxyear* and *minyear* and missing a record, a manual fill for that author-affiliation pair is created.

Affiliation	y2005	y2006	y2007	y2008	y2009	y2010	minyear	maxyear
A	1	1	.	.	.	.	2005	2006
B	.	1	1	F	F	1	2006	2010
C	.	1	F	F	1	1	2006	2010

Appendix: Constructing Author-Affiliation-Year Record

Step 3: For any given year, if there is record of any other affiliation for said author, then the filled records are removed.

Affiliation	y2005	y2006	y2007	y2008	y2009	y2010	minyear	maxyear
A	1	1	.	.	.	.	2005	2006
B	.	1	1	F	.	1	2006	2010
C	.	1	.	F	1	1	2006	2010

Appendix: Constructing Author-Affiliation-Year Record

Step 4: For any given year that only has filled records, only the filled record that is closest to the last actual record is kept.

Affiliation	y2005	y2006	y2007	y2008	y2009	y2010	minyear	maxyear
A	1	1	.	.	.	.	2005	2006
B	.	1	1	F	.	1	2006	2010
C	.	1	.	.	1	1	2006	2010

Appendix: Constructing Author-Affiliation-Year Record

These steps will yield the long-form of the data:

year	affiliation
2005	A
2006	A
2006	B
2006	C
2007	B
2008	B
2009	C
2010	B
2010	C

Distribution of Papers Across Days, Months, and Years

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
1	81.7	82.7	84.0	81.9	80.2	78.8	77.7	77.0	78.3	78.4	80.9	82.8	92.6	94.7
2-10	5.0	4.7	4.1	4.7	5.8	6.1	6.7	7.3	6.4	6.1	5.4	5.4	2.0	1.7
11-20	6.5	5.9	5.8	6.2	6.3	6.8	7.6	7.1	7.5	7.2	6.6	5.7	2.3	1.8
21-31	6.8	6.6	6.1	7.2	7.7	8.4	8.0	8.6	7.8	8.3	7.1	6.0	3.0	1.7
Jan.	61.8	64.8	65.5	60.4	61.1	59.4	54.1	55.2	59.2	56.8	61.0	68.0	85.0	89.6
Feb.-Jun.	18.2	15.7	15.6	17.8	17.0	17.3	20.0	20.9	19.0	19.5	17.9	18.3	7.6	6.7
Jul.-Sep.	10.4	9.5	8.9	10.7	10.7	11.3	13.2	12.2	10.9	11.9	11.1	7.3	4.0	2.6
Oct.-Dec.	9.6	10.1	10.0	11.1	11.2	12.0	12.7	11.7	10.8	11.7	10.0	6.3	3.3	1.0