

Gender and Coauthorship at the Federal Reserve Board

Deepa D. Datta Robert J. Vigfusson

Federal Reserve Board

October 2019

The views expressed are those of the authors and do not necessarily reflect the views of anyone else in the Federal Reserve System.

Overview

- ▶ We gathered data on Federal Reserve Board (FRB) economists since 2003.
- ▶ Combining this information with data on all FRB working papers, we study coauthorship patterns among FRB economists.
- ▶ **Main Finding:** The observed distribution of coauthorship groupings across genders differs from predictions based on random assignment.
- ▶ **Potential Harm:** Gender bias in coauthorship may lower research productivity among women, as measured by the number of working papers per person.
 - ▶ Research output is an important input into promotion decisions.
 - ▶ Lower productivity may hamper career progression.

Why examine gender patterns in coauthorship?

Coauthors generally boost research productivity.

Barriers to coauthorship are barriers to productivity.

- ▶ Economics has trended towards more coauthorship.
 - ▶ For FRB working papers, solo-authored papers are down from 57% of papers in earlier decades to 26% in recent decade.
 - ▶ 3-authored papers are up from 6 to 27%, papers with 4 or more authors are up from 2 to 10%.

Why examine gender patterns in coauthorship?

Coauthors generally boost research productivity.

Barriers to coauthorship are barriers to productivity.

- ▶ Economics has trended towards more coauthorship.
 - ▶ For FRB working papers, solo-authored papers are down from 57% of papers in earlier decades to 26% in recent decade.
 - ▶ 3-authored papers are up from 6 to 27%, papers with 4 or more authors are up from 2 to 10%.

Coauthorship patterns can help us understand gender patterns in research productivity and career progression.

- ▶ Divergence from random assignment may indicate bias or lack of inclusivity in coauthorships, or more broadly in the profession.

Data

Our data is from public-facing sources.

1. **Papers** scraped from REPEC.

- ▶ Universe of FEDS/IFDPs, covering 3,170 papers published between 1971 and 2018.
- ▶ Titles, WP year, author names (1,857 unique authors).
- ▶ Downloads, abstract views, JEL codes, publication, year.

Data

Our data is from public-facing sources.

1. **Papers** scraped from REPEC.

- ▶ Universe of FEDS/IFDPs, covering 3,170 papers published between 1971 and 2018.
- ▶ Titles, WP year, author names (1,857 unique authors).
- ▶ Downloads, abstract views, JEL codes, publication, year.

2. **FRB economists** scraped from the Board's public website, as archived by the Wayback Machine.

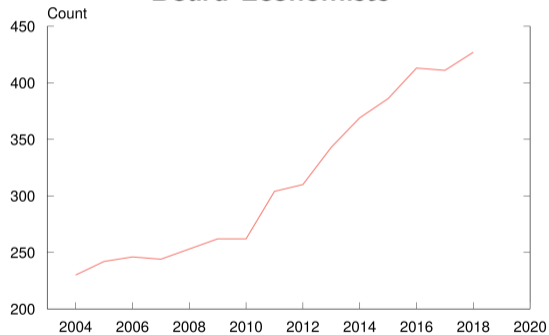
- ▶ FRB website lists economist names, Ph.D. year, position, section, and division.
- ▶ Wayback machine provides 181 captures between March 2003 and January 2019, tracking economists' years of service and promotions over time.
- ▶ We track 663 FRB economists. 513 of these are also authors of working papers.

Gender Tagging

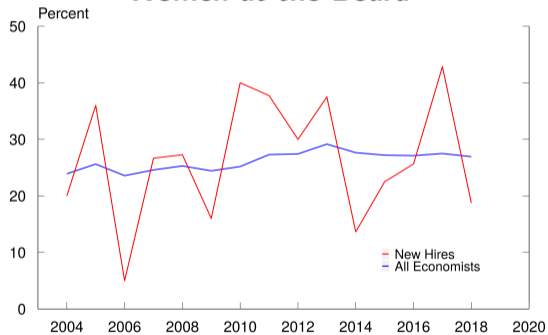
- ▶ We gender tag the 2,007 individuals:
 - ▶ using existing data sets of economists from Hengel (2017) and Chari and Goldsmith-Pinkham (2017),
 - ▶ using the Tang, et al. (2011) gender dictionary,
 - ▶ and manually through personal knowledge and web searches.
- ▶ 587 are identified by the existing data sets, 769 identified using the Tang dictionary, and 651 are manually identified.
- ▶ Overall, we have 451 women and 1,556 men.

Potential Coauthors - Women at the Board

Board Economists



Women at the Board



Potential Coauthors - Women in Economics

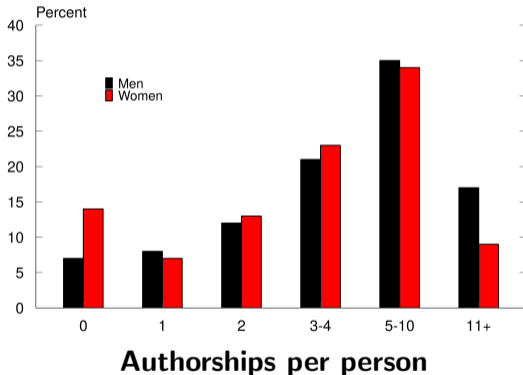
- ▶ In academia: Women accounted for 23.5% of tenured and tenure-track faculty in economics in 2016 (Bayer and Rouse, 2016).
 - ▶ About 31% of U.S. Ph.D.s in economics in 2014 were granted to women.
- ▶ In macro/finance fields: Women accounted for about 25% of dissertations in 2015 (based on data in Lundberg and Stearns, 2018).
- ▶ In FRB working papers: Women accounted for 22% of the external coauthors on papers posted between 2004 and 2017.

Distribution of Authorships

The median number of authorships is 3 for both men and women.
The mean is a bit higher for men (4.9) than for women (3.9).

Distribution of Authorships

The median number of authorships is 3 for both men and women.
The mean is a bit higher for men (4.9) than for women (3.9).



Note: Sample restricted to FRB economists employed in 2017 with at least 3 years of service (n=307).

Interpreting Coauthorship Statistics

38% of two-authored papers have at least one woman author.

Is that good? We'll use a model to figure it out.

- ▶ Take as given the number of authors on a particular paper.
- ▶ Assume authors are chosen at random, with a fixed probability of an author being a woman.
- ▶ How frequently should we observe different gender combinations under this model? How does this compare to the actual data?

Random Assignment - Two Author Case

Probability that both authors are women ($X = 2$)

$$\Pr(X = 2 | n = 2, p = f) = f^2$$

Random Assignment - Two Author Case

Probability that both authors are women ($X = 2$)

$$\Pr(X = 2 | n = 2, p = f) = f^2$$

Probability that both authors are men ($X = 0$)

$$\Pr(X = 0 | n = 2, p = f) = (1 - f)(1 - f)$$

Random Assignment - Two Author Case

Probability that both authors are women ($X = 2$)

$$\Pr(X = 2|n = 2, p = f) = f^2$$

Probability that both authors are men ($X = 0$)

$$\Pr(X = 0|n = 2, p = f) = (1 - f)(1 - f)$$

Probability that exactly one author is a woman ($X = 1$)

$$\Pr(X = 1|n = 2, p = f) = \mathbf{2f(1 - f)}$$

Random Assignment - Two Author Case

Probability that both authors are women ($X = 2$)

$$\Pr(X = 2|n = 2, p = f) = f^2$$

Probability that both authors are men ($X = 0$)

$$\Pr(X = 0|n = 2, p = f) = (1 - f)(1 - f)$$

Probability that exactly one author is a woman ($X = 1$)

$$\Pr(X = 1|n = 2, p = f) = 2f(1 - f)$$

For an n -authored paper, probability of having k women authors

$$\Pr(X = k|n, p = f) = \binom{n}{k} f^k (1 - f)^{n-k}$$

Predicted Coauthorship

	2004-2010 Predicted	2011-2017 Predicted
1 Author		
Man	75%	72%
Woman	25%	28%
2 Authors		
Men	57%	52%
Mixed	37%	40%
Women	6%	8%
3+ Authors		
Men	41%	36%
Majority Men	42%	43%
Mixed	4%	5%
Majority Women	12%	15%
Women	1%	2%
FRB Women Economists	24.6%	27.6%

Predicted vs Actual Coauthorship

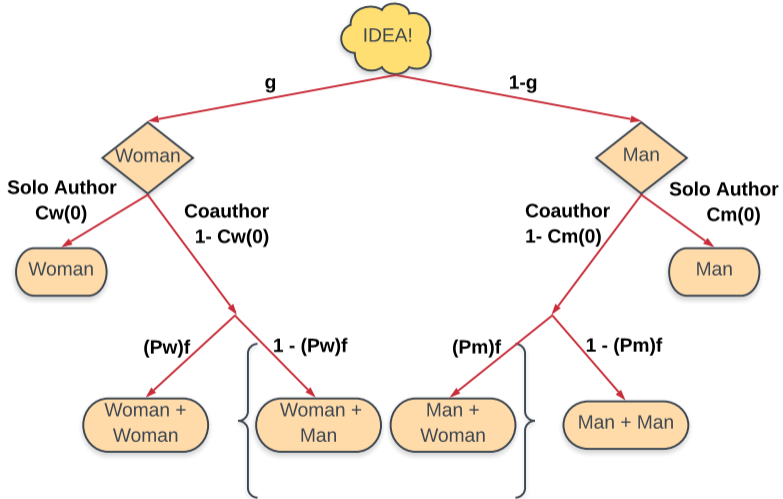
	2004-2010		2011-2017	
	Predicted	Actual	Predicted	Actual
1 Author				
Man	75%	81%	72%	76%
Woman	25%	19%	28%	24%
2 Authors				
Men	57%	71%	52%	62%
Mixed	37%	25%	40%	30%
Women	6%	5%	8%	8%
3+ Authors				
Men	41%	49%	36%	45%
Majority Men	42%	40%	43%	33%
Mixed	4%	2%	5%	4%
Majority Women	12%	7%	15%	11%
Women	1%	2%	2%	6%
FRB Women Economists		24.6%		27.6%

A Richer Model

Observed pattern differs from random assignment.

Next step is to use a richer, more-detailed model to structure our thinking about the observed pattern.

A Richer Model - Two Author Case



A Richer Model - Two Author Case

Outcome	Probability
W writes alone	$g c_w(0)$
W writes with W	$g c_w(1) p_w f$
W and M write together	$g c_w(1)(1 - p_w f)$ $+ (1 - g) c_m(1) p_m f$
M writes with M	$(1 - g) c_m(1)(1 - p_m f)$
M writes alone	$(1 - g) c_m(0)$

$p_w \in [0, \frac{1}{f}]$, $p_m \in [0, \frac{1}{f}]$:

- ▶ $p_i = 0 \implies$ never choose a woman coauthor
- ▶ $p_i < 1 \implies$ probability of matching with a woman is less than the frequency of potential women coauthors
- ▶ $p_i = \frac{1}{f} \implies$ only choose a woman coauthor

A Richer Model - Four Author Case

- ▶ $z(n, k)$, the observed frequency of n -authored papers with k women authors.
- ▶ Model-predicted frequencies of different gender groupings depends on $\theta \equiv \{f, g, p_w, p_m, c_w(0), c_w(1), c_w(2), c_m(0), c_m(1), c_m(2)\}$

Define a loss function, $L(\theta)$, as the sum of squared differences between observed and predicted frequencies:

$$L(\theta) \equiv \sum_{n=1}^4 \sum_{k=0}^n (z(n, k) - \Pr(X = k, n|\theta))^2$$

Choose $\hat{\theta}$ to minimize the loss function, given observed frequencies:

$$\hat{\theta} = \operatorname{argmin} L(\theta)$$

Estimated Model Parameters

Sample	f	g	p_w	p_m	$c_w(0)$	$c_m(0)$
2004-2010	0.246	0.240	0.59	0.47	0.28	0.37

- ▶ $g \approx f$: Women initiate projects in proportion to their representation in the population.
- ▶ $p_w < 1, p_m < 1$: Consistent with women being underrepresented as coauthors.
- ▶ $p_m < p_w$: Men seem less likely than women to match with a woman coauthor.
- ▶ $c_m(0) > c_w(0)$: Men seem more likely than women to write solo-authored papers.

Estimated Model Parameters

Sample	f	g	p_w	p_m	$c_w(0)$	$c_m(0)$
2004-2010	0.246	0.240	0.59	0.47	0.28	0.37
2011-2017	0.276	0.210	1.65	0.82	0.30	0.25

- ▶ $f \uparrow, g \downarrow$: Low g may reflect challenges women face initiating projects.
- ▶ p_w and p_m both increased.
- ▶ $p_w > 1$: Women are more likely to coauthor with each other than would be suggested by chance.
- ▶ $p_w > 1, p_m < 1$: Both men and women are making choices that limit the number of gender mixed teams.

Some Alternative Estimates

Sample	f	g	p_w	p_m	$c_w(0)$	$c_m(0)$	W-Authorships
2011-2017	0.276	0.210	1.65	0.82	0.30	0.25	0.241
Alt 1	0.276	0.210	1.65	1.00	0.30	0.25	0.262
Alt 2	0.276	0.276	1.26	0.61	0.24	0.27	0.244
Alt 3	0.276	0.276	1.00	1.00	0.24	0.27	0.274

1. Setting $p_m = 1 \implies$ W-Authorships improve but do not reach f .
2. Setting $g = f \implies$ lower p_w and p_m are required to match the data.
3. With $g = f$, setting $p_w = p_m = 1$ results in W-Authorships $\approx f$.

Discussion

- ▶ The observed distribution of coauthorship groupings across genders differs from predictions based on random assignment.
- ▶ Gender bias in coauthorship selections may result in lower observed productivity among women, as measured by the number of working papers per person.
 - ▶ Important for understanding career progression, as research output is often cited as an important input into promotion and award decisions.
- ▶ To the extent that there is learning by doing in research, barriers to finding coauthors in early years may result in lower productivity throughout a person's career.

Writing and Career Progression

Years 1-3			Years 4-6		
	Men	Women		Men	Women
1. Writes	72%	52%	1. Writes	71%	74%
			2. Doesn't	13%	15%
			3. Exits	16%	11%
2. Doesn't	17%	33%	1. Writes	42%	24%
			2. Doesn't	46%	59%
			3. Exits	13%	18%
3. Exits	11%	15%			

Note: Includes economists starting after March 2003 and before January 2012.

Observed Patterns in Experience Pairings

We calculate the observed distribution of coauthor experience, given author experience and gender.

Coauthor Age	All			Women			Men		
	Author Age			Author Age			Author Age		
	R	M	S	R	M	S	R	M	S
R	0.42	0.13	0.12	0.35	0.23	0.08	0.44	0.10	0.12
M	0.21	0.44	0.21	0.35	0.48	0.16	0.16	0.43	0.22
S	0.38	0.43	0.68	0.30	0.29	0.76	0.40	0.47	0.66

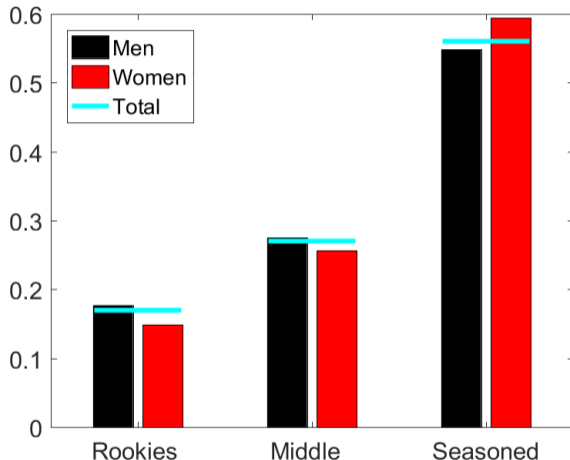
Categories: Rookies: 3 years or less since Ph.D., Middle: 4-7 years, Seasoned: 8+ years.

Observed Patterns in Experience Pairings

Coauthor	All			Women			Men		
	R	M	S	R	M	S	R	M	S
R	0.42	0.13	0.12	0.35	0.23	0.08	0.44	0.10	0.12
M	0.21	0.44	0.21	0.35	0.48	0.16	0.16	0.43	0.22
S	0.38	0.43	0.68	0.30	0.29	0.76	0.40	0.47	0.66

- ▶ **Diagonals** are large. Highly likely to coauthor with others in same age cohort.
- ▶ Again, we turn to random sampling for a benchmark comparison.

Experience Distribution of FRB Economists



Fraction of economists between 2004-17, by time since Ph.D. graduation.

Categories: Rookies: 3 years or less since Ph.D., Middle: 4-7 years, Seasoned: 8+ years.

Predicted vs Actual Experience Pairings

Population	All			Women			Men		
	R	M	S	R	M	S	R	M	S
R=0.17	0.42	0.13	0.12	0.35	0.23	0.08	0.44	0.10	0.12
M=0.27	0.21	0.44	0.21	0.35	0.48	0.16	0.16	0.43	0.22
S=0.56	0.38	0.43	0.68	0.30	0.29	0.76	0.40	0.47	0.66

- ▶ Relative to random assignment, the big takeaway is that Rookies and Middles are much more likely to write with their own Cohort.
- ▶ Rookie Women coauthor less with Seasoned than Rookie Men.
- ▶ Relative to Middle Men, Middle Women write more with Rookies and less with Seasoned.

Conclusion

- ▶ Our results focus primarily on gender groupings.
 - ▶ Assortative matching can occur along other characteristics, including age, common language, nationality, university, physical location, or interests.
- ▶ In general, our data can provide insight into inclusivity in coauthorships and the economics profession overall.
- ▶ Large literature showing that diverse groups have better outcomes in a variety of settings.
- ▶ As such, reducing barriers to diverse coauthorships, perhaps especially for rookies, may improve outcomes.
- ▶ In future work, we plan to explore this hypothesis further through examination of data on paper downloads, abstract views, and time to publication.