Self-Selection and the Diminishing Returns of Research

Lorenz Ekerdt, *Stony Brook*

Kai-Jie Wu, *Penn State*

May 2025

Researcher productivity has fallen dramatically in the U.S.

- Massive increase in # of researchers, but flat/declining TFP growth
- Used by literature to infer, e.g.,
  - diminishing returns in R&D: "ideas getting harder to find" (Bloom et al, 2020)
  - worsening misallocation in R&D (Çelik, 2023; Lehr, 2025; Fernández-Villaverde, 2025)
  - lower R&D spillovers (Dyévre, 2025)

This paper: quality-adjusted researcher input w.r.t. composition effects

- Composition effect:
  - expansion of R&D share in laborforce
  - self-selection in researchers' ability

  $\Big\}$ $\Longrightarrow$ decline in R&D productivity

- Estimate Roy-like researcher supply model
- **Finding:** adjustment cut **half** of increase in researcher input

Researchers are getting harder to find too!

Outline

1. Motivating evidence

2. Model

3. Targeted moments

4. Estimation

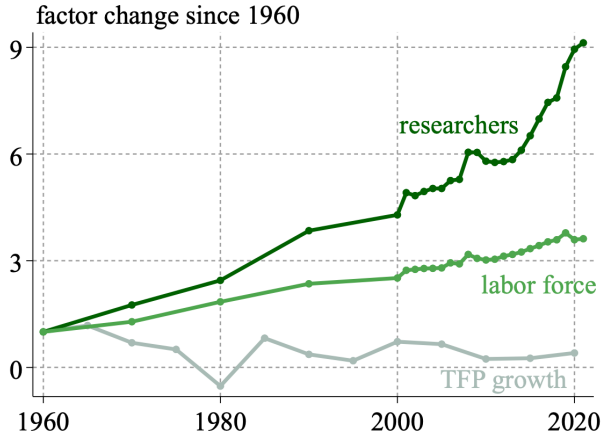
5. Implications for semi-endogenous growth models

6. Discussion and limitations

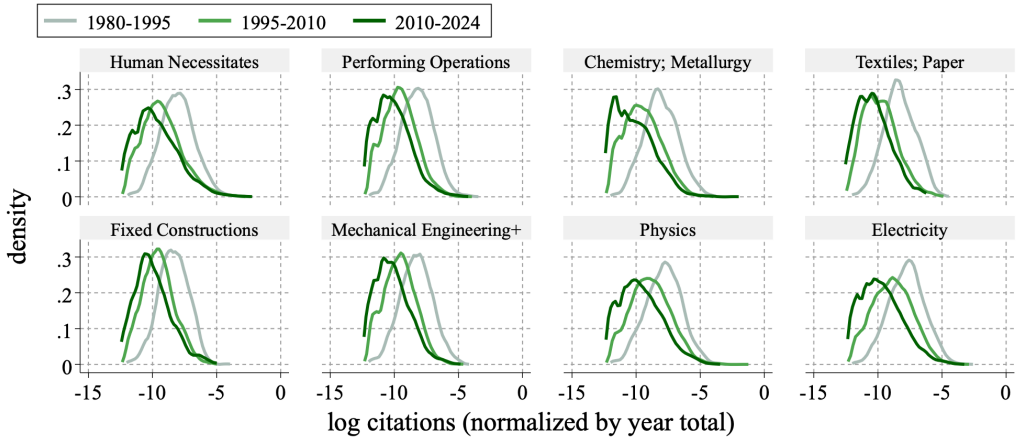
Outline

1. Motivating evidence
2. Model
3. Targeted moments
4. Estimation
5. Implications for semi-endogenous growth models
6. Discussion and limitations

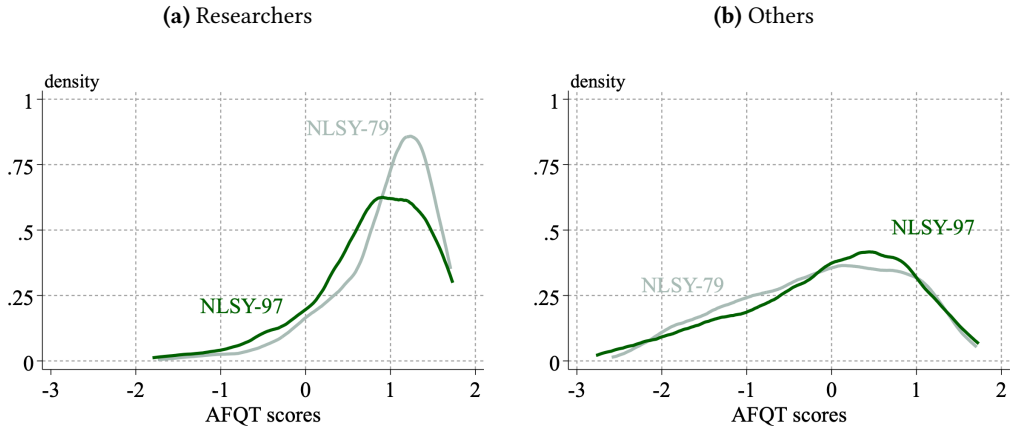
Declining researcher productivity and R&D sector expansion



Evidence 1: Lower patent productivity for later-cohort innovators



Evidence 2: Lower test scores for later-cohort researchers



Notes: standardized AFTQ scores taken from Altonji, Bharadwaj, Lange (2009)

Outline

1. Motivating evidence
2. Model
3. Targeted moments
4. Estimation
5. Implications for semi-endogenous growth models
6. Discussion and limitations

Roy-like self-selected researcher supply

- Three sectors j
  - Low-skill production (N)
  - High-skill production (H), requires college education
  - Researchers (R), requires college education
- Workers i heterogeneous in:
  - sector-neutral gain from college z_{Ci}
  - sector-specific high-skill ability (z_{Ri}, z_{Hi})
- Labor markets competitive with log wage rate w_{jt} for each $j \in \{N, H, R\}$

- Education choice $s \in \{0, 1\}$ with costs μ_C

$$u_i = \max_{s \in \{0,1\}} \{s \cdot (z_{Ci} + u_{Ci} - \mu_C)\}$$

- Sectoral choice $j \in \{H, R\}$ with costs μ_j

$$u_{Ci} = \mathbb{E}\left(\max_{j \in \{R,H\}} \{z_{ji} + w_j - \mu_j\}\right)$$

- **Composition effects:**

$$\overline{z}_{R,t} = \underbrace{\mathbb{E}\left(z_{Ci} \mid s_i = 1\right)}_{\text{in education}} + \underbrace{\mathbb{E}\left(z_{Ri} \mid s_i = 1, j_i = R\right)}_{\text{in sectoral choice}}$$

Outline

1. Motivating evidence
2. Model
3. Targeted moments
4. Estimation
5. Implications for semi-endogenous growth models
6. Discussion and limitations

Data: worker-level

National Survey of College Graduates (NSCG)

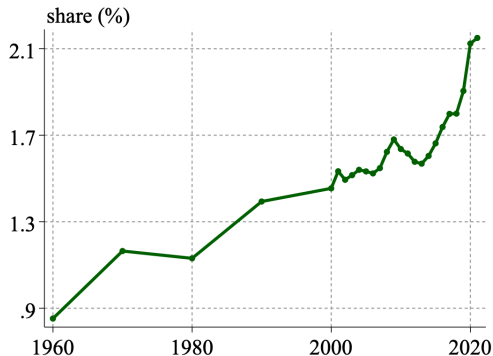
- education, occupation, (hourly) earnings, etc.
- **primary work activity ⟶ R&D**
- 2003, 2010, 2013, 2015, 2017, 2019, 2021

Decennial Census and American Community Survey (DC+ACS)

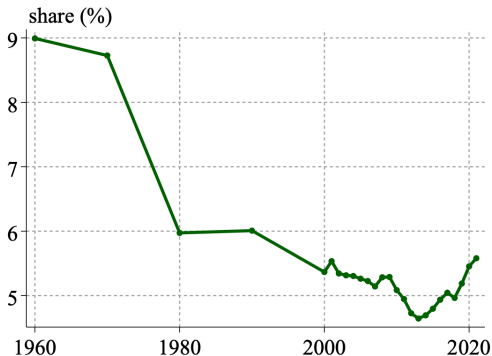
- education, occupation, (hourly) earnings, etc.
- nationally representative
- 1960-1990 (decennial), 2000-2021 (annual)

Researcher share increases overall, but decreases among college grads

(a) Researcher share in labor force

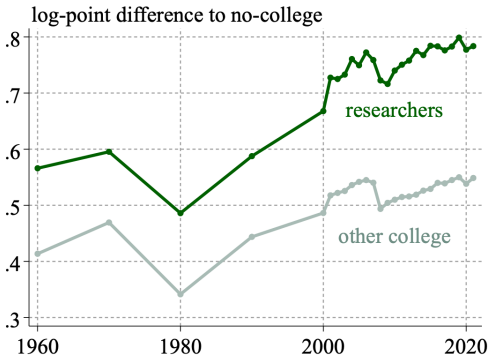


(b) Researcher share among college grads

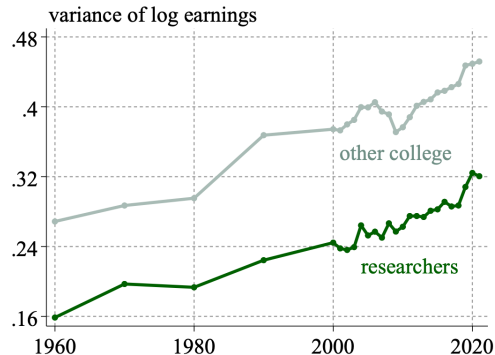


Researchers' and other college grads' earnings increasingly disperse over time

(a) Average log earning relative to N



(b) Variance of log earnings

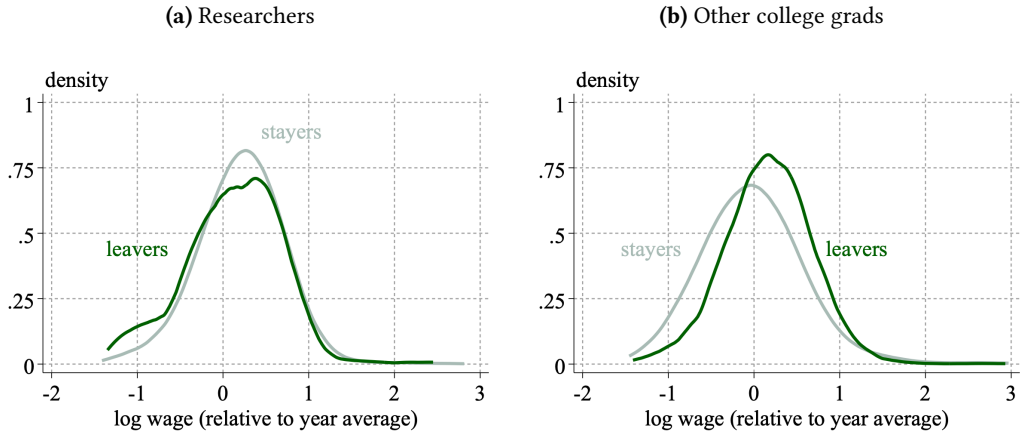


NSCG Panel: compare earnings between sector stayers vs. movers

Recall: key determinant of selection is marginal workers vs. conditional average

- Assume **movers $\longleftrightarrow$ close to margin**
  - movers < stayers $\longrightarrow$ positive self-selection into sector
  - stayers < movers $\longrightarrow$ negative self-selection into sector
- NSCG: same respondent identifiers in the 2010, 2013, 2015 surveys

Panel moments from 2010-2013 NSCG



Outline

1. Motivating evidence
2. Model
3. Targeted moments
4. Estimation
5. Implications for semi-endogenous growth models
6. Discussion and limitations

Estimate the model by indirect inference

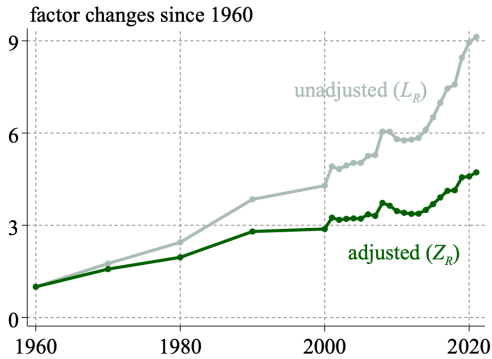
Parameters to be estimated:

- ability distribution parameterized by joint normal (fixed over time)
- sectoral wages and costs in 1960
- changes of sectoral wages and costs between 1960 and 2021

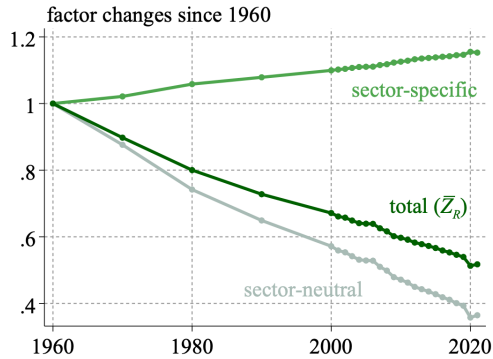
Targeted moments:

- 1960 U.S. labor market moments
- 1960-2021 changes of sector shares and wage dispersion
- longitudinal moments in NSCG panel

Estimation results: adjust research input down by half



(a) Total researchers: L_R vs. Z_R



(b) Average researcher ability

External validation for self-selection in college education

1. Counterfactual change in college wage premium w/o self-selection

- Carneiro and Lee (2011): 30% higher
- Our model: **40% higher**

2. Difference in average latent ability btw college and non-college workers

- Hendricks and Schoellman (2014): 1.44 to 3.75 std dev depending on specification
- Heckman et al. (2018): 2 std dev
- Our model: **1.59** std dev in 1960

Outline

1. Motivating evidence
2. Model
3. Targeted moments
4. Estimation
5. Implications for semi-endogenous growth models
6. Discussion and limitations

Double long-run growth

- Jones' (1995) idea production: $\dot{A}_t = A_t^{\phi} Z_{R,t}$
- ϕ determines long-run growth: $g_A^{\text{BGP}} \propto g_L / (1-\phi)$
- How to pin down ϕ?

$$\underbrace{\Delta \ln\left(\dot{A}_t / A_t\right)}_{\approx 0} = (\phi - 1) \underbrace{\Delta \ln A_t}_{>0} + \underbrace{\Delta \ln Z_{R,t}}_{\gg 0} \implies 1 - \hat{\phi} = \frac{\Delta \ln Z_{R,t}}{\Delta \ln A_t}$$

- Quality adjustment $\longrightarrow$ half $(1-\hat{\phi})$ $\longrightarrow$ **double** g_A^{BGP}

Outline

1. Motivating evidence
2. Model
3. Targeted moments
4. Estimation
5. Implications for semi-endogenous growth models
6. Discussion and limitations

Key: infer self-selection from changing dispersion in R and H

Question: Do all changes in wage dispersion reflect self-selection?

Potential challenges:

1. Changing misallocation across demographic groups

2. Factors that changes dispersion in all sectors, including N

Appendix

Relation with Bloom et. al (2020)'s measure

- Measure effective number of researchers by:

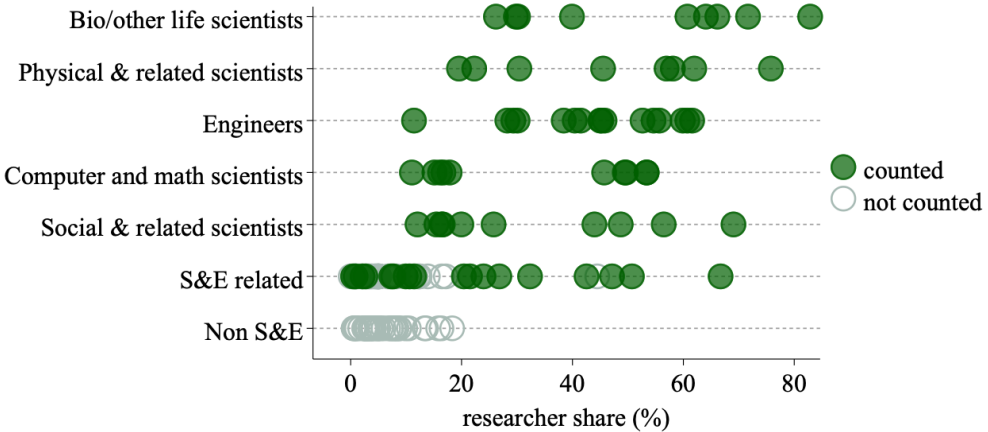
$$\hat{Z}_{Rt} = \frac{\text{R\&D expenditure}}{\text{college grads average earnings}} = \left(\frac{W_{Rt}}{\overline{W}_{Ct}}\right) Z_{Rt}$$

- The average wage of college graduates is

$$\overline{W}_{Ct} = \left(\frac{L_{Rt}}{L_{Ct}}\right) W_{Rt}\overline{Z}_{Rt} + \left(\frac{L_{Ht}}{L_{Ct}}\right) W_{Ht}\overline{Z}_{Ht},$$

- $\hat{Z}_{Rt}$ **at best captures** L_{Rt} **if** $W_{Rt} = W_{Ht}$ **and** $\overline{Z}_{Rt} = \overline{Z}_{Ht}$

Share of researchers in each NSCG occupation



Estimated ability distribution

Parameter	Description	Value
	Latent ability distribution	
σ_C	variance of ability gained from college	1.1128
σ_R	variance of ability in sector R	0.2281
σ_H	variance of ability in sector H	0.1692
ρ	correlation between abilities in sector R and H	0.9661

Parameter	Description	Value
	Sectoral wages and costs in 1960	
w_R	log efficiency wage of sector R	−1.7561
w_H	log efficiency wage of sector H	−1.5476
μ_C	log costs of college education	−0.0972
$\mu_R - \mu_H$	log relative costs of becoming a researcher	−0.1039
	Changes in relative returns from 1960 to 2021	
$\Delta(\tilde{w}_R - \tilde{w}_H)$	relative log wage-to-cost ratio	−0.0195
$\Delta(u_C - \mu_C)$	mean net log return of college	1.1353

Exactly-matched moments

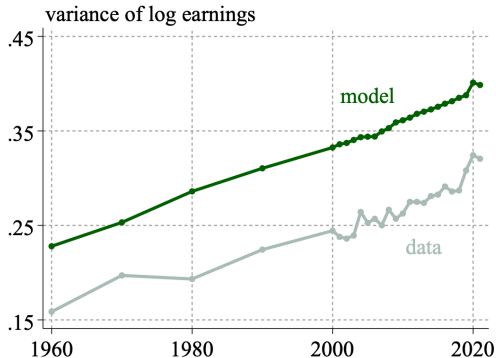
Moment	Notation	Data = Model
Initial values in 1960		
share of college grads	S_C	0.0948
share of researchers among college grads	$S_{R\|C}$	0.0899
relative mean log earnings of researchers	E_R	0.5660
relative mean log earnings of other college grads	E_H	0.4136
Changes from 1960 to 2021		
share of college grads	ΔS_C	0.2905
share of researchers among college grads	$\Delta S_{R\|C}$	−0.0341
Longitudinal moments in 2010-2015 NSCG		
share of movers in R	$S_{R\to H}$	0.2032
share of movers in H	$S_{H\to R}$	0.0115

Numerically-approached moments: data vs. model

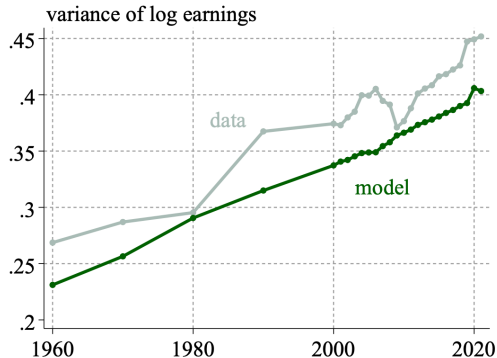
Moment	Notation	Data	Model
Changes in earnings dispersion from 1960 to 2021			
researchers	ΔV_R	0.1618	0.1702
other college grads	ΔV_H	0.1832	0.1729
Longitudinal moments in 2010-2015 NSCG			
mean log earnings, leavers minus stayers in R	$E_R^{R\to H}$	-0.0846	-0.0862
mean log earnings, leavers minus stayers in H	$E_H^{H\to R}$	0.1829	0.1854

Earnings dispersion among college graduates, model vs. data

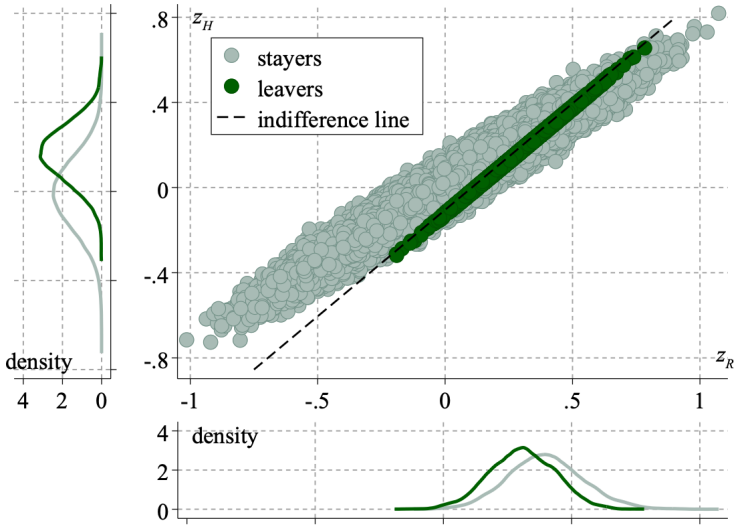
(a) Researchers V_R



(b) Other college grads V_H



How panel moments identify distribution of sectoral-specific abilities



Magnitude of self-selection

Positive and large self-selection in college education

- mainly governed by ΔV_H and ΔV_R in data
- 1pp increase in the $S_C \longrightarrow$ 0.19% decrease in $\mathbb{E}(z_{Ci}|s_i = 1)$

Positive but small self-selection from H to R

- mainly governed by longitudinal moments
- 1pp increase in the $S_{R|C} \longrightarrow$ 0.03% decrease in $\mathbb{E}(z_{Ri}|s_i = 1, j_i = R)$

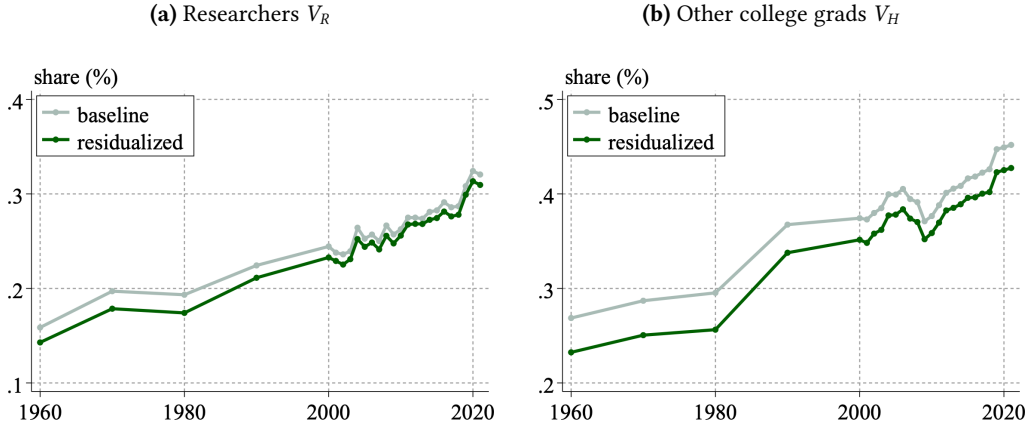
Clarifying remarks for permanent diminishing returns

- Different from labor diminishing returns in the form of $\lambda < 1$ where

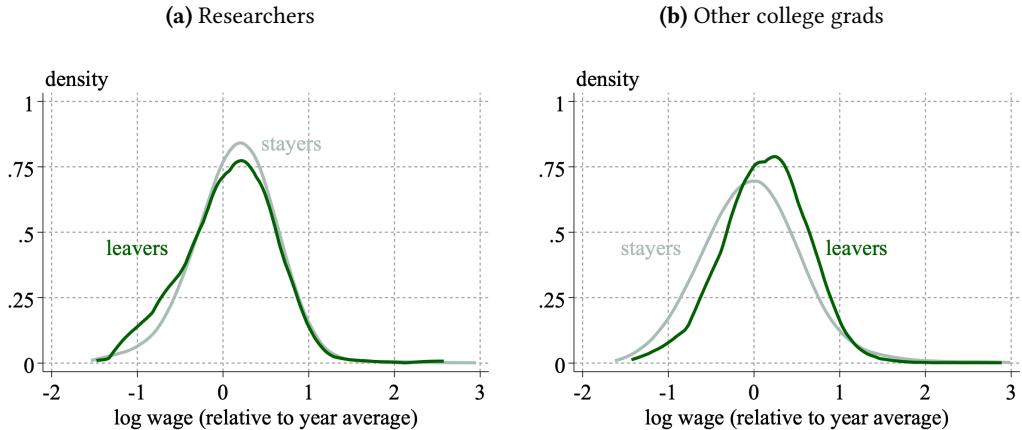
$$\dot{A} = A^{\phi} {L_R}^{\lambda}$$

- diminishing returns w.r.t. **stock** vs. **share** of researchers
- λ is permanent diminishing return
- isomorphic to ϕ in predicting long-run growth, $g \propto \lambda n/(1-\phi)$

Earnings dispersion among college graduates, baseline vs. residualized



Residualized earnings: stayers v.s movers



Changes researcher ability, allowing for common external factors

