Self-Selection and the Diminishing Returns of Research

Lorenz Ekerdt (U.S. Census Bureau)[†]

Kai-Jie Wu (Penn State)

January 2025

[†] *Any views expressed are those of the authors and not those of the U.S. Census Bureau.*

R&D productivity has fallen dramatically in the U.S.

- R&D is the engine of long-term growth (Romer, 1990)
- Massive increase in R&D , but flat/declining TFP growth
- Seen by literature as the **permanent diminishing returns** of R&D
  - "ideas getting harder to find" (Bloom et al, 2020)
  - key for long-run prediction in semi-endogenous growth models

This paper: part of the decline is transitory

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

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

- **Transitional** because it vanishes when R&D share ceases to rise
- Literature attribute all decline to permanent $\Longrightarrow$ understate long-run growth

Researchers are getting harder to find too!

Roadmap to quantify the transitional composition effect

1. **Construct growth model with Roy-like selection into research**

   key: employment share of researchers increasing $\longrightarrow$ composition effects

2. **Document facts of U.S. labor market of researchers**

   microdata on workers' education, occupation, work activity, and earnings

3. **Estimate model using data**

   infer positive selection into research, mostly through college education

4. **Separate transitional and permanent diminishing returns**

   predicted long-run growth rate nearly doubles

Related Literature

1. Growth and permanent diminishing returns in R&D

   Jones (1995, 2002); Kortum (1997); Bloom et al; (2020), Peters (2022)

2. Self-Selection in higher education

   Carneiro and Lee (2011); Hendricks, Schoellman (2014); Heckman, Humphries, Veramendi (2018)

3. Estimating Roy models

   Roy (1951); Heckman, Sedlacek (1985); Heckman, Honoré (1990); French, Taber (2011)

4. Aggregate & policy implications of self-selected labor allocation

   Lagakos, Waugh (2013); Young (2014); Hsieh et al. (2019); Akcigit, Pearce, Prato (2024)

Outline

1. Model: permanent vs. transitory diminishing returns

2. Empirical findings
   - Trends in U.S. researchers' labor market
   - Evidence for self-selection from NSCG panel

3. Estimation

4. Revisiting the permanent diminishing returns

Jones' (1995) growth model with self-selected labor supply

Three types of labor:

- Low-skill production (N)
- High-skill production (H), requires college education
- Researchers (R), requires college education

Two production sectors: $Z_{j,t}$ efficiency unit of labor input

- Aggregate production equivalence: $Y_t = A_t G(Z_{H,t}, Z_{N,t})$
- Research: $\dot{A}_t = A_t^{\phi} Z_{R,t}$

Labor markets are competitive with log wage rate w_{jt} for each $j \in \{N, H, R\}$

Self-selection in education and sectoral labor supply

L_t measure of workers, heterogeneous in:

- sector-neutral gain from college $e^{z_{Ci}}$
- sector-specific high-skill ability $(e^{z_{Ri}}, e^{z_{Hi}})$

Education choice $s \in \{0, 1\}$ with opportunity costs μ_C (Willis, Rosen, 1987)

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

Sectoral choice $j \in \{H, R\}$ with non-pecuniary returns μ_j

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

Permanent diminishing returns to research: $\phi \leq 1$

Idea production: $\dot{A}_t = A_t^{\phi} Z_{R,t}$

- $\phi > 0$: standing on the shoulders of giants
- $\phi < 0$: fishing out of a pond

ϕ determines predicted long-run growth rate:

$$g_A^{\text{BGP}} = \left(\frac{1}{1-\phi}\right) g_{z_R}^{\text{BGP}} = \left(\frac{1}{1-\phi}\right) g_L,$$

Transitional diminishing returns (composition effect)

- Define $z^m_{C,t}$ and $z^m_{R,t}(z_H)$ as ability of workers **on the margin**
- average researcher ability: $\overline{z}^R_t = \mathbb{E}\left[z_{Ci} + z_{Ri} \mid z_{Ci} \geq z^m_{C,t}, z_{Ri} \geq z^m_{R,t}(z_{Hi})\right]$
- **Composition effect:** assume z_{Ci} independent of (z_{Ri}, z_{Hi})

$$\dot{\overline{z}}^R_t = \underbrace{\left(\dot{\tilde{w}}_{R,t} + \dot{\tilde{w}}_{H,t}\right)\left[z^m_{C,t} - \mathbb{E}\left(z_{Ci} \mid z_{Ci} > z^m_{C,t}\right)\right]}_{\text{selection into college}}$$

$$+ \underbrace{\left(\dot{\tilde{w}}_{R,t} - \dot{\tilde{w}}_{H,t}\right)\left(\mathbb{E}\left[z^m_{R,t}(z_{Hi})\right] - \mathbb{E}\left[z_{Ri} \mid z_{Ri} > z^m_{R,t}(z_{Hi})\right]\right)}_{\text{selection into research among college grads}}$$

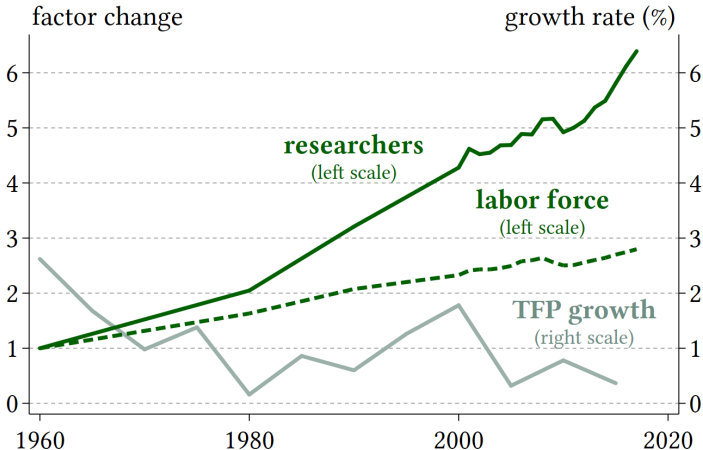
- Sectoral expansion lowers average ability if
  - marginal worker has lower ability than conditional average
  - call self-selection “positive” in this case
- **Transitional:** vanishes when sectoral composition reaches constant on BGP

Observed decline in research productivity: permanent or transitional?

Rearrange idea production function:

$$\underbrace{\Delta_k \ln\left(\dot{A}_t/A_t\right)}_{\approx 0} - \underbrace{\Delta_k \ln L_{R,t}}_{\gg 0} = (\phi - 1) \underbrace{\Delta_k \ln A_t}_{>0} + \Delta_k \ln \overline{Z}_{R,t}$$

- Literature attribute all decline to permanent: $\phi \ll 1$ assuming $\Delta_k \ln \overline{Z}_{R,t} = 0$
- We argue: research expansion cause composition effect: $\Delta_k \ln \overline{Z}_{R,t} \neq 0$
  - Strong enough to significantly change estimated ϕ?
  - Need to estimate the model appropriately


factor change
growth rate (%)
6
5
4
3
2
1
0
researchers
(left scale)
labor force
(left scale)
TFP growth
(right scale)
1960
1980
2000
2020

Outline

1. Model: permanent vs. transitory diminishing returns

2. Empirical findings
   - Trends in U.S. researchers' labor market
   - Evidence for self-selection from NSCG panel

3. Estimation

4. Revisiting the permanent diminishing returns

Data

National Survey of College Graduates (NSCG)

- education, occupation, earnings, (age, gender, race, etc.)
- **primary work activity ⟶ R&D**
- 1993, 2003, 2010, 2013, 2015, 2017

Decennial Census and American Community Survey (DC+ACS)

- education, occupation, earnings, (age, gender, race, etc.)
- nationally representative
- 1960-1990 (decennial), 2000-2017 (annual)

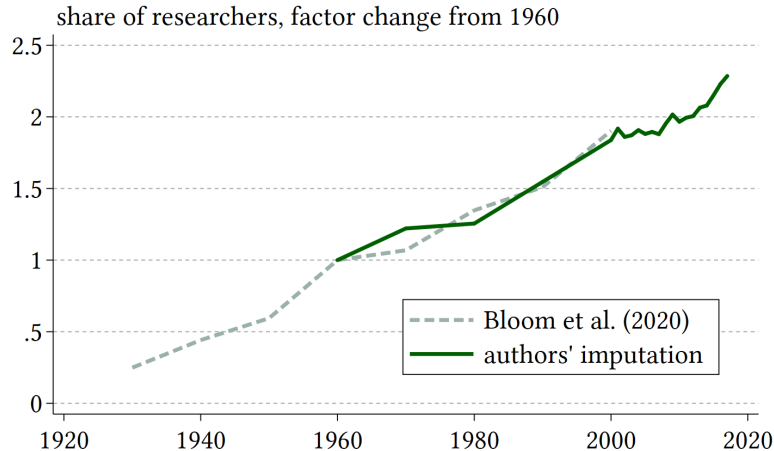
Outline

1. Model: permanent vs. transitory diminishing returns
2. Empirical findings
   - Trends in U.S. researchers' labor market
   - Evidence for self-selection from NSCG panel
3. Estimation
4. Revisiting the permanent diminishing returns

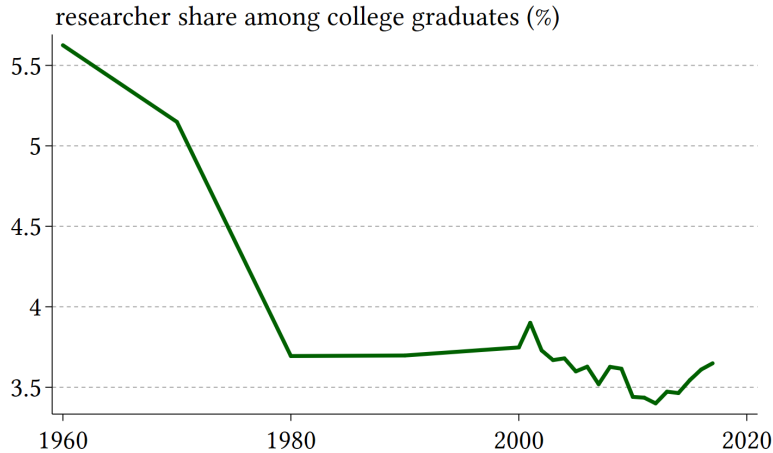
Constructing aggregate trends of U.S. researcher's labor market

1. Identify researchers in the NSCG
2. Construct occupational researcher shares in NSCG
3. Impute to DC+ACS to construct moments on sectoral shares and earnings

Researcher share increases overall

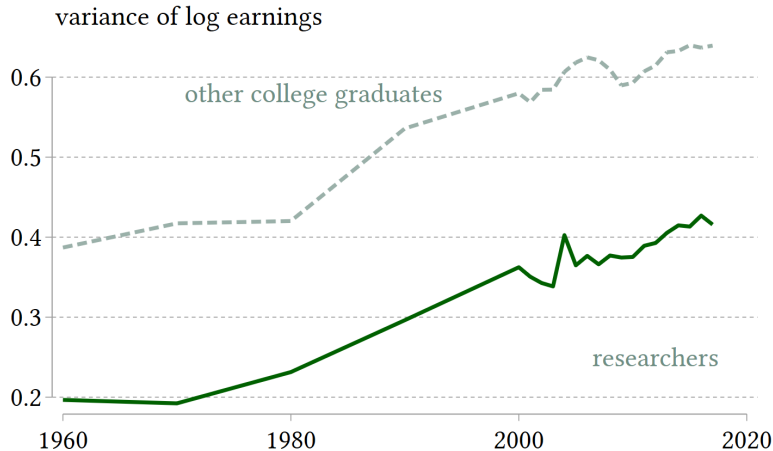


But decreases among college grads



Researchers' and other college grads' relative earnings increase over time

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



Recap

Findings:

- increasing share of researchers ($S_R = L_R/L$) and other college grads ($S_H = L_H/L$)
- increasing rel. earnings of researchers ($E_R - E_N$) and other college grads ($E_H - E_N$)
- increasing earning dispersion for researchers (V_R) and other college grads (V_H)

Implications:

- (positive) selection from no-college (N) to the other two (R & H)
- but unclear for selection between researchers (R) and other college grads (H)

Outline

1. Model: permanent vs. transitory diminishing returns
2. Empirical findings
   - Trends in U.S. researchers' labor market
   - Evidence for self-selection from NSCG panel
3. Estimation
4. Revisiting the permanent diminishing returns

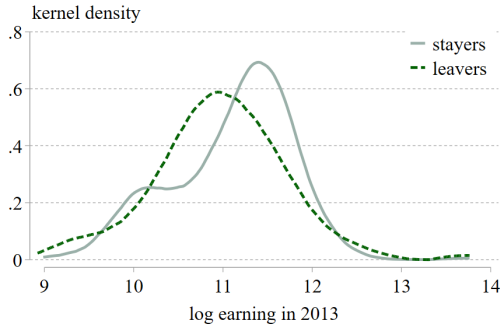
Compare earnings between sector stayers vs. movers

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

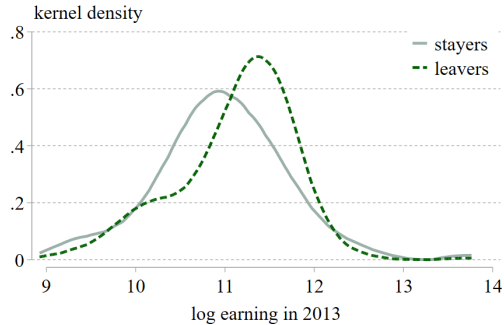
- Assume **movers $\longleftrightarrow$ close to margin**
  - movers < stayers $\longrightarrow$ positive self-selection into sector
  - stayers < movers $\longrightarrow$ negative self-selection into sector
- NSCG: same respondent identifiers in the 2010, 2013, 2015 surveys

Panel moments from 2013-2015 NSCG

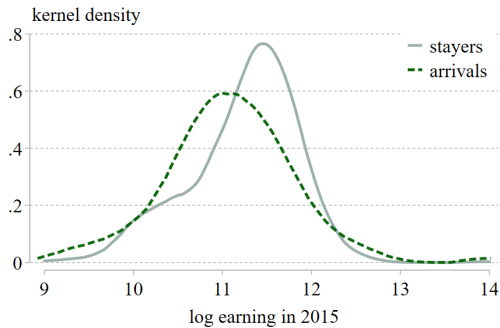
(a) Researchers in 2013



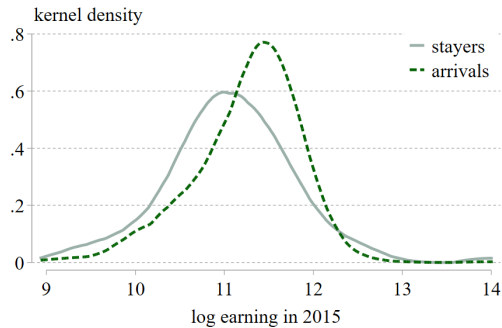
(b) Other college grads in 2013



(a) Researchers in 2015



(b) Other college grads in 2015



Recap

Findings:

- among researchers, **stayers earn more than movers**
- among other college grads, **stayers earn less than movers**

Implications:

- positive self-selected into research
- negative self-selected into non-research other college grads

Outline

1. Model: permanent vs. transitory diminishing returns
2. Empirical findings
   - Trends in U.S. researchers' labor market
   - Evidence for self-selection from NSCG panel
3. Estimation
4. Revisiting the permanent diminishing returns

Estimate the model by indirect inference

Parameters to be estimated:

- latent ability distribution: parameterized by joint normal
- sectoral wages and costs in 1960
- changes of sectoral wages and costs between 1960-2017

Targeted moments:

- 1960 U.S. labor market moments
- 1960-2017 changes of sector shares and wage dispersion
- longitudinal moments in 2013-2015 NSCG panel

Summarize estimation results

Positive and large self-selection in college education

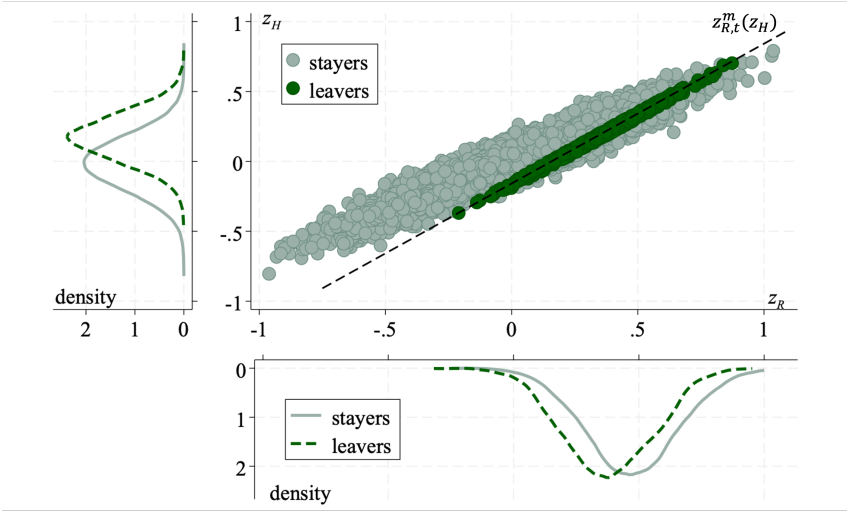
- result of large estimated σ_C, which mainly governed by ΔV_H and ΔV_R in data
- 1pp increase in the S_C is associated w/ 0.24% decrease in $\mathbb{E}(z_{Ci}|s_i = 1)$

Positive but small self-selection from H to R

- mainly governed by longitudinal moments
- 1pp increase in the $S_{R|C}$ is associated w/ 0.01% decrease in $\mathbb{E}(z_{Ri}|s_i = 1, j_i = R)$

$\longrightarrow$ average researcher ability decline by 48% from 1960 to 2017

How panel moments identify distribution of sectoral-specific abilities



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: **32% 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: **1.86** (in 2017) to **1.93** (in 1960) std dev

Outline

1. Model: permanent vs. transitory diminishing returns
2. Empirical findings
   - Trends in U.S. researchers' labor market
   - Evidence for self-selection from NSCG panel
3. Estimation
4. Revisiting the permanent diminishing returns

Separating transitory diminishing returns from permanent

- Recall permanent diminishing return:

$$\phi = 1 - \frac{\Delta \ln L_R + \Delta \ln \overline{Z}_R}{\Delta \ln A}$$

- E.g. TFP growth rate = 1.5%, population growth rate = 1%
  - assume $\Delta \ln \overline{Z}_R = 0 \implies \phi = -0.72$ and $g_A^{\text{BGP}} = 0.58\%$
  - our estimated $\Delta \ln \overline{Z}_R \implies \phi = 0.07$ and $g_A^{\text{BGP}} = 1.08\%$ (1.85 times higher)
- **fishing out $\longrightarrow$ standing on the shoulders of giants**

Conclusion

Part of decline in research productivity is transitional composition effect

1. Document facts of U.S. labor market of researchers
2. Quantify transitory diminishing returns by Roy model
3. Revisit long-run prediction in semi-endogenous growth models

Appendix

Clarifying remarks

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

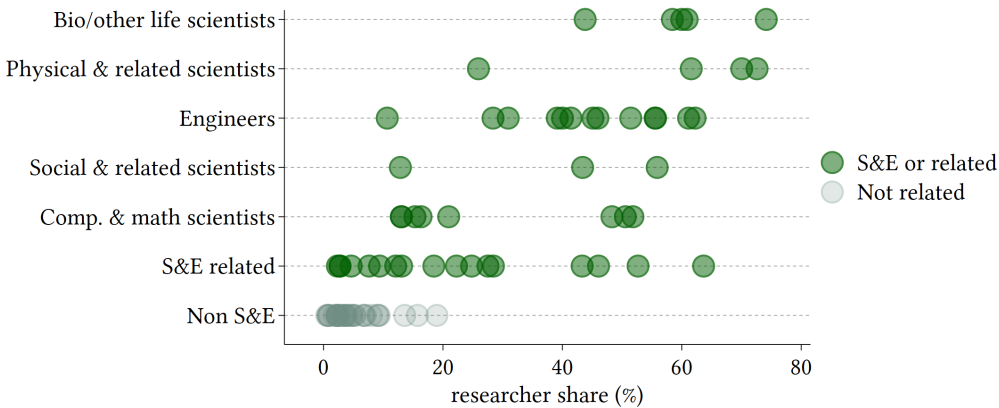
$$\dot{A} = A^{\phi} {L_R}^{\lambda}$$

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

Identifying researchers in NSCG

- Label a worker as a researcher if:
  - primary work activity (PWA) = “research” or “development”
  - occupation ∈ “scientists and engineers (S&E)” or “S&E related”
- Define **share of researchers** for each occupation j in NSCG:

$$s_j^{\text{NSCG}} = \left[\frac{\sum_{i \in I_j} \mathbb{1}(\text{PWA}_i = \text{R\&D})}{L_j} \right] \times \mathbb{1}\left(j \in \left\{ \text{S\&E or related} \right\} \right)$$


Bio/other life scientists
Physical & related scientists
Engineers
Social & related scientists
Comp. & math scientists
S&E related
Non S&E
0
20
40
60
80
researcher share (%)
S&E or related
Not related

For each occupation j in NSCG,

- Share of researchers' log earnings

$$s_{e,j}^{\text{NSCG}} = \left[\frac{\sum_{i \in I_j} e_i \cdot \mathbb{1}(\text{PWA}_i = \text{R\&D})}{\sum_{i \in I_j} e_i} \right] \times \mathbb{1}\left(j \in \left\{ \text{S\&E or related} \right\} \right).$$

- Share of researchers' squared log earnings

$$s_{e2,j}^{\text{NSCG}} = \left[\frac{\sum_{i \in I_j} e_i^2 \cdot \mathbb{1}(\text{PWA}_i = \text{R\&D})}{\sum_{i \in I_j} e_i^2} \right] \times \mathbb{1}\left(j \in \left\{ \text{S\&E or related} \right\} \right).$$

Impute occupational researcher shares from NSCG to DC/ACS

- Number of researchers

$$L_R = \sum_{j \in \mathcal{J}} L_{Cj} \cdot s_j^{NSCG}$$

where L_{Cj} is the number of college-educated workers in occupation j

- Mean log earnings of researchers

$$E_R = \frac{1}{L_R}\left[\sum_{j}\left(\sum_{i\in C_j} e_i\right) s_{e,j}^{NSCG}\right].$$

- Variance log earnings of researchers

$$V_R = \frac{1}{L_R}\left[\sum_{j\in J}\left(\sum_{i\in C_j} {e_i}^2\right) s_{e2,j}^{NSCG}\right] - {E_R}^2,$$

- same way define L_H, E_H, and V_H for other college grads.

Parameterize ability distribution by joint normal

- sector-neutral college learning ability:

$$z_{Ci} \sim N(0, {\sigma_C}^2)$$

- sector-specific ability (assume independent from z_{Ci})

$$\begin{pmatrix} z_{H,i} \\ z_{R,i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_H^2 & \rho\sigma_H\sigma_R \\ \rho\sigma_H\sigma_R & \sigma_R^2 \end{pmatrix} \right]$$

Remark: cannot separate mean and wage → capture composition effect only

- **Goal:** $\overline{Z}_R = \mathbb{E}\left[\exp(z_{Ci} + z_{Ri}) \mid s_i = 1, j_i = R\right]$
- **Key:** self-selection governed by ability distribution
  - Educational choice:

$$\mathbb{E}\left(z_{Ci} \mid s_i = 1\right) = \left(\frac{\sigma_C}{S_C}\right) \phi \left(\frac{u_C - \mu_C}{\sigma_C}\right)$$

  - Sectoral choice: let $\theta = \sqrt{\sigma_R^2 + \sigma_H^2 - 2\rho\sigma_R\sigma_H}$

$$\mathbb{E}\left(z_{Ri} \mid s_i = 1, j_i = R\right) = \left(\frac{\sigma_R^2 - \rho\sigma_R\sigma_H}{\theta \; S_{R|C}}\right) \phi \left[\frac{(w_R - \mu_R) - (w_H - \mu_H)}{\theta}\right]$$

Discussion of distributional assumptions

1. Fully parametric: no instrument covers support for long-term implication
2. Other common distributions strongly restrict self-selection (e.g., extreme value)
3. Independence restricting education to be positively selected
   - wealth of evidence in literature: Hendricks, Schoellman (2014), Heckman et al (2018)
   - but do not restrict strength selection: largely determined by σ_C^2

Constructing longitudinal moments in the model

No gross flows in our model, instead:

1. Calculate change in returns that generates mover shares equal to data
2. Calculate moments of log earnings for movers vs. stayers generated above

Estimated ability distribution

Parameter	Description	Value
	Latent ability distribution	
σ_C	variance of ability gained from college	1.335
σ_R	variance of ability in sector R	0.258
σ_H	variance of ability in sector H	0.210
ρ	correlation between abilities in sector R and H	0.961

Parameter	Description	Value
	Sectoral wages and costs in 1960	
w_R	log efficiency wage of sector R	-1.869
w_H	log efficiency wage of sector H	-1.928
μ_C	log costs of college education	-0.252
$\mu_H - \mu_R$	log costs of becoming a researcher	0.187
	Changes in relative wages and costs from 1960 to 2017	
$\Delta(\tilde{w}_R - \tilde{w}_H)$	relative log wage-to-cost ratio	-0.144
$\Delta(u_C - \mu_C)$	mean net return of college	0.445

Exactly-matched moments

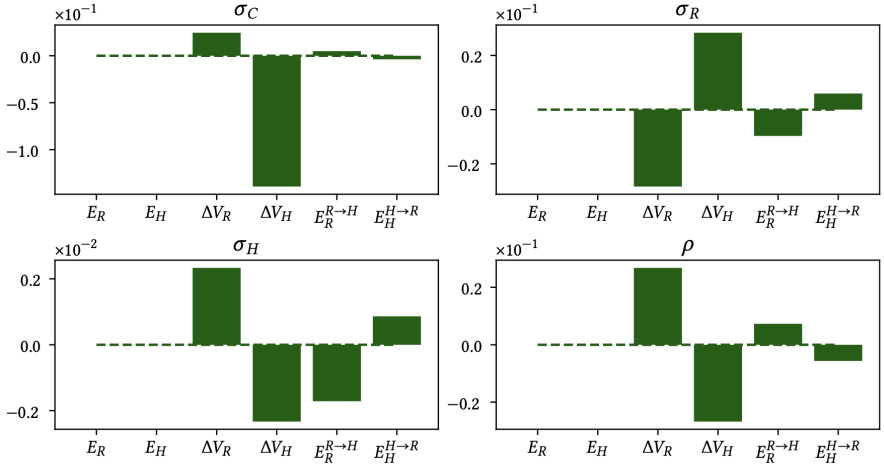
Moment	Notation	Data
Initial sectoral shares in 1960		
share of college grads (%)	S_C	10.4
share of researchers among college grads (%)	$S_{R\|C}$	5.63
Changes in sectoral shares from 1960 to 2017		
share of college grads (pp)	ΔS_C	26.5
share of researchers among college grads (pp)	$\Delta S_{R\|C}$	-1.98
Longitudinal moments in 2010-2015 NSCG		
share of movers in R (%)	$S_{R\to H}$	24.0
share of movers in H (%)	$S_{H\to R}$	3.5

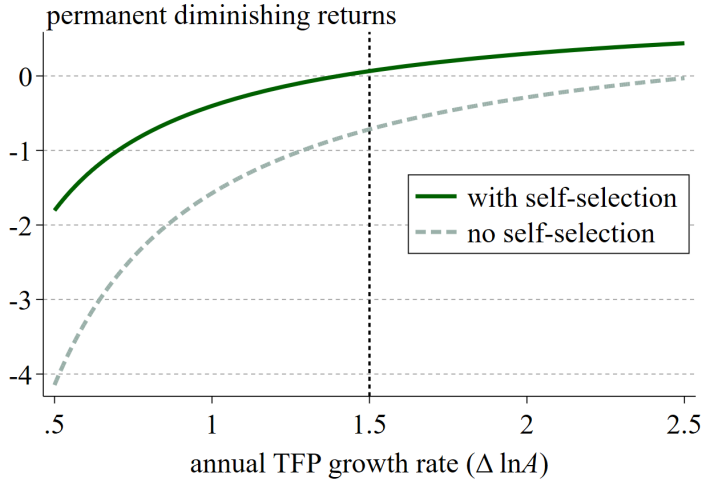
Numerically-approached moments: data vs. model

Moment	Notation	Data	Model
Initial mean log earnings in 1960			
relative mean log earnings of researchers	E_R	0.801	0.801
relative mean log earnings of other college grads	E_H	0.372	0.372
Changes in earnings dispersion from 1960 to 2017			
researchers	ΔV_R	0.212	0.223
other college grads	ΔV_H	0.243	0.224
Longitudinal moments in 2010-2015 NSCG			
mean log earnings, leavers minus stayers in R	$E_R^{R \to H}$	-0.080	-0.079
mean log earnings, leavers minus stayers in H	$E_H^{H \to R}$	0.180	0.172

Sensitivity of estimated parameters w.r.t moments

Andrews, Gentzkow, and Shapiro (2017)




permanent diminishing returns
0
-1
-2
-3
-4
.5
1
1.5
2
2.5
annual TFP growth rate (Δ ln$A$)
with self-selection
no self-selection