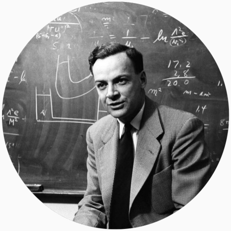
Competing for Scientific Talent: Industry vs. Academia

Justine Boudou

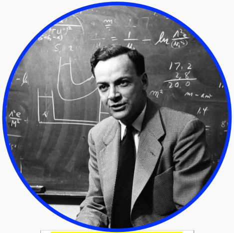
04/26/2024

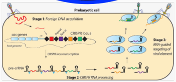
Harvard Business School

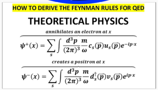
Scientists drive innovation and economic growth



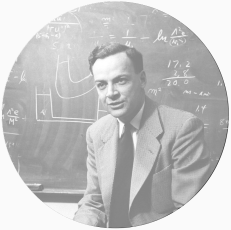
Scientists can create value in Academia

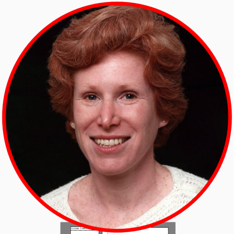


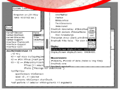


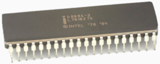


Scientists can create value in Industry



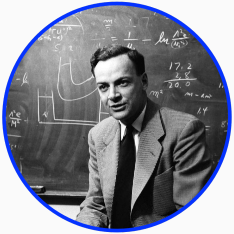






Markets don't necessarily direct R&D to areas of highest value

Academia



Industry

How much friction is there in the allocation of scientists across sectors?

Controversies about the (mis)allocation of scientists

Universities' AI Talent Poached by Tech Giants

Google, Baidu, Facebook lure away top computer science researchers responsible for training students

Controversies about the (mis)allocation of scientists

Universities' AI Talent Poached by Tech Giants

Google, Baidu, Facebook lure away top computer science researchers responsible for training students

Why economists are flocking to Silicon Valley

And why big tech wants them

Controversies about the (mis)allocation of scientists

Universities' AI Talent Poached by Tech Giants

Google, Baidu, Facebook lure away top computer science researchers responsible for training students

Why economists are flocking to Silicon Valley

And why big tech wants them

INNOVATIONS

Why Silicon Valley needs humanities PhDs

Controversies about the (mis)allocation of scientists

Universities' AI Talent Poached by Tech Giants

Google, Baidu, Facebook lure away top computer science researchers responsible for training students

Why economists are flocking to Silicon Valley

And why big tech wants them

INNOVATIONS

Why Silicon Valley needs humanities PhDs

REPORT: Careers Outside of Academia are Richly Rewarding for PhD Physicists

Why do we care today?

- Knowledge-base economy: intellectual capital is a key driver of economic growth (Romer, 1990; Bell et al., 2017)

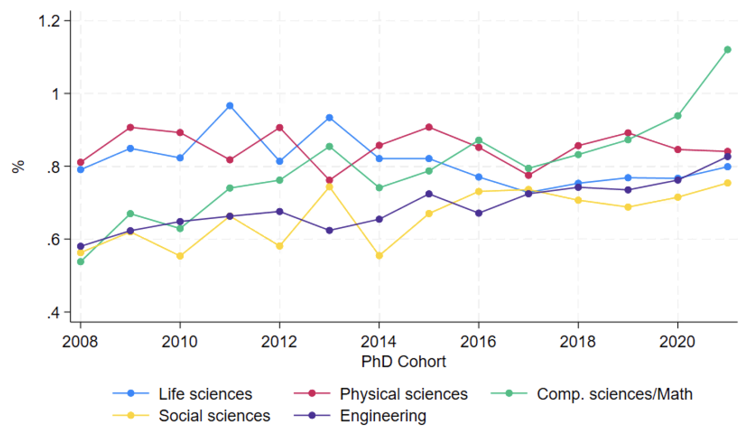
Why do we care today?

- Knowledge-base economy: intellectual capital is a key driver of economic growth (Romer, 1990; Bell et al., 2017)
- Competition for talent: firms need talent to maintain competitiveness ; universities need talent for prestige and reputation

Why do we care today?

- Knowledge-base economy: intellectual capital is a key driver of economic growth (Romer, 1990; Bell et al., 2017)
- Competition for talent: firms need talent to maintain competitiveness ; universities need talent for prestige and reputation
- Increasing earnings gap at graduation between Industry and Academia

Why do we care today?



Earnings gap at graduation in Industry and Academia, by cohort and field

Research Question

- **How socially efficient is the allocation of scientists across sectors?**

Research Question

- **How socially efficient is the allocation of scientists across sectors?**
- **How do scientists perform in Industry vs Academia?**

Research Question

- **How socially efficient is the allocation of scientists across sectors?**
- **How do scientists perform in Industry vs Academia?**

To answer this question we need:

1. Measures of private and social value

Research Question

- **How socially efficient is the allocation of scientists across sectors?**
- **How do scientists perform in Industry vs Academia?**

To answer this question we need:

1. Measures of private and social value
2. Identification strategy that accounts for selection and heterogeneity

This paper

- Measures of private and social value: focus on **earnings** and **publications**
    - Private value: **earnings** + non-pecuniary returns
    - Social value: Private value + Externalities (**publications**, ...)

This paper

- Measures of private and social value: focus on **earnings** and **publications**
  - Private value: **earnings** + non-pecuniary returns
  - Social value: Private value + Externalities (**publications**, ...)
  - Administrative dataset (SED/SDR) that matches demographics and earnings information about doctorate holders who graduated between 1970 and 2013 to their publication output

This paper

- Measures of private and social value: focus on **earnings** and **publications**
  - Private value: **earnings** + non-pecuniary returns
  - Social value: Private value + Externalities (**publications**, ...)
  - Administrative dataset (SED/SDR) that matches demographics and earnings information about doctorate holders who graduated between 1970 and 2013 to their publication output
- Identification strategy that accounts for **selection** and **heterogeneity**
  - Model of selection
  - Within department variation
  - Instrumental variable (IV)
  - Estimation of Marginal Treatment Effects (MTE)

This paper

- Measures of private and social value: focus on **earnings** and **publications**
  - Private value: **earnings** + non-pecuniary returns
  - Social value: Private value + Externalities (**publications**, ...)
  - Administrative dataset (SED/SDR) that matches demographics and earnings information about doctorate holders who graduated between 1970 and 2013 to their publication output
- Identification strategy that accounts for **selection** and **heterogeneity**
  - Model of selection
  - Within department variation
  - Instrumental variable (IV)
  - Estimation of Marginal Treatment Effects (MTE)
- (Mis)Allocation - *Preliminary results*

Preliminary Findings

- Joining Industry (vs Academia) leads to:
  - A large **increase** in **lifetime earnings**, driven by the early-stage of the career
  - A large **decrease** in the overall **stock of publications**

Preliminary Findings

- Joining Industry (vs Academia) leads to:
    - A large **increase** in **lifetime earnings**, driven by the early-stage of the career
    - A large **decrease** in the overall **stock of publications**

- I find substantial **heterogeneity** and **selection** (ATT>ATE)

Sector joined	β^W
Industry	85%
Academia	2%

Preliminary Findings

- Joining Industry (vs Academia) leads to:
  - A large **increase** in **lifetime earnings**, driven by the early-stage of the career
  - A large **decrease** in the overall **stock of publications**

- I find substantial **heterogeneity** and **selection** (ATT>ATE)

Sector joined	β^W	β^{Pub}
Industry	85%	-60%
Academia	2%	-81%

Preliminary Findings

- Joining Industry (vs Academia) leads to:
    - A large **increase** in **lifetime earnings**, driven by the early-stage of the career
    - A large **decrease** in the overall **stock of publications**
- I find substantial **heterogeneity** and **selection** (ATT>ATE)

Sector joined	β^W	β^{Pub}
Industry	85%	-60%
Academia	2%	-81%

- The earnings premium in Industry appears relatively stable over time
- The publications gap increased in the 90s and started to decrease again in the mid 2000s

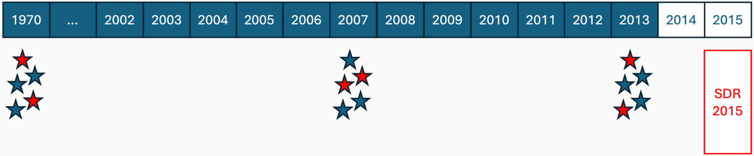
Contributions

- Literature on the role of Industry and Academia in the production of knowledge and innovation
  - Difference in institutional norms: (Dasgupta and David, 1994; Aghion et al., 2008; Stephan, 2012; Furman and Stern, 2011; Kahn and Ginther, 2017; Stephan, 1996; Stephan et al., 2004; Agarwal and Ohyama, 2013)
  - Differences in individuals' preferences: (Stern, 2004; Roach and Sauermann, 2010, 2023, 2024; Sauermann and Roach, 2012; Sauermann and Stephan, 2013)
- Literature on the relationship between talent, innovation and economic growth
  - Frictions related to the extensive margin: Bell et al. (2017), Akcigit et al. (2020), Celik (2023)
  - Frictions related to the intensive margin: Akcigit and Goldschlag (2023)
- Literature on resource allocation within organizations and its consequences for performance
  - Literature on general and firm-specific human capital: Coff and Kryscynski (2011)
  - Selection: Shu (2016)

Data

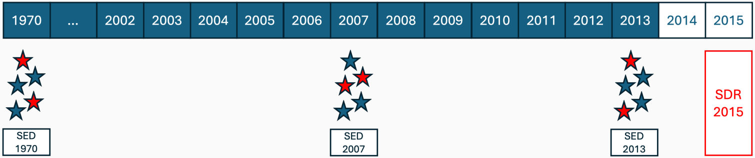
Data Sources

- **Survey of Doctorate Recipients (SDR 2015)**: *(during the career)*
  - Across cohorts, longitudinal
  - **Earnings (flow)**, Patents (flow, sub-sample)
- **Survey of Doctorate Recipients Bibliometric Research Data**:
  - SDR 2015 matched with Web of Science
  - **Publication output (stock)** from PhD graduation to 2017



Data Sources

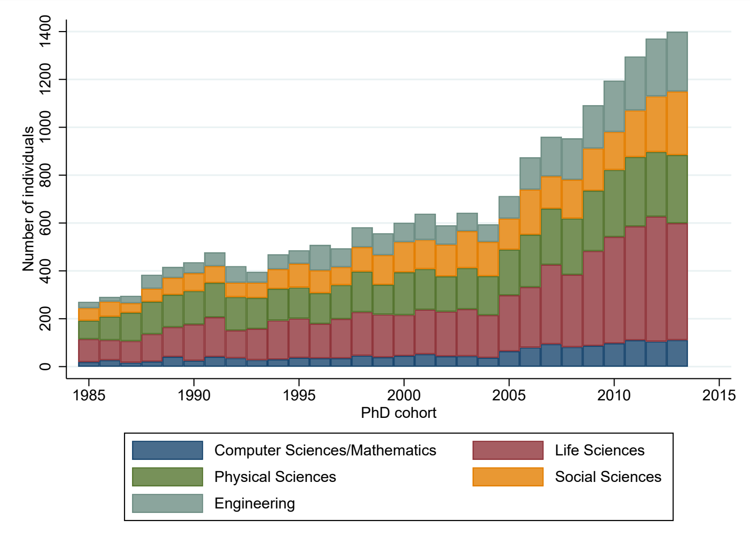
- **Survey of Doctorate Recipients (SDR 2015)**: *(during the career)*
  - Across cohorts, longitudinal
  - **Earnings (flow)**, Patents (flow, sub-sample)
- **Survey of Doctorate Recipients Bibliometric Research Data**:
  - SDR 2015 matched with Web of Science
  - **Publication output (stock)** from PhD graduation to 2017
- **Survey of Earned Doctorates (SED)**: *(at graduation)*
  - Within cohort
  - Sector joined at graduation (+ earnings at start)



Sample and main variables

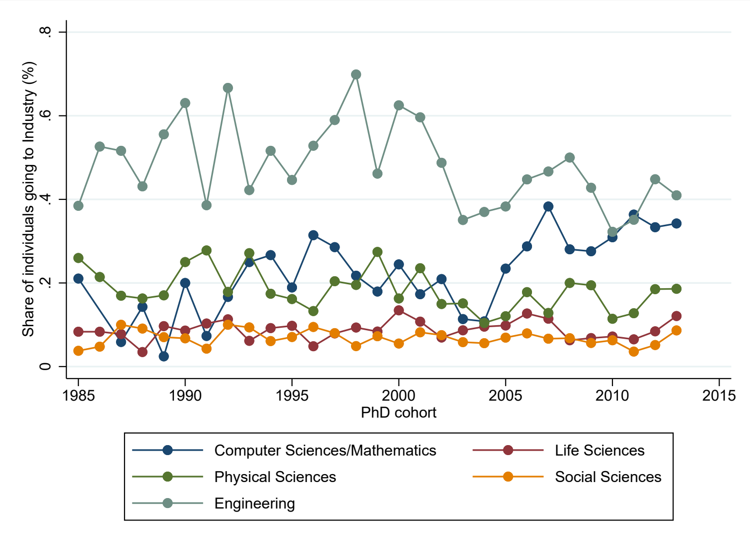
- **21,406 individuals**
  - Ind. with PhD cohort $\geq$ 1970, with definite postgraduate commitment at the time of survey and start their career in the U.S.
- **Treatment**:
  - *Starting* in Industry vs Academia (incl. Government/Postdoc)
- **Outcomes**:
  - Earnings: total earned income before deductions in the previous year (2015 US dollars)
  - Publications: from PhD graduation to the time of survey (2017)

Supply

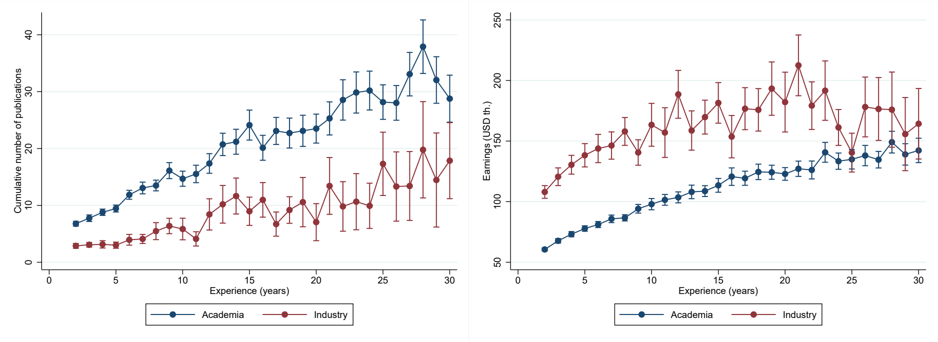


57 PhD majors, 5 fields (CS/Maths, Life Sc., Phys. Sc., Social Sc., Eng.)

Sorting at graduation, by PhD field



Raw differences - Earnings and Publications



(a) Cumulative number of publications

(b) Earnings (in 2015 USD th.)

Empirical Framework and Results

Overview

1. 'Control' for ability: within department variation

⬇

2. Selection at the individual level (LATE)

⬇

3. Selection at the individual level + Heterogeneity (MTE)

⬇

4. (Mis)allocation

Overview

1. ‘Control’ for ability: within department variation

⬇

2. Selection at the individual level (LATE)

⬇

3. Selection at the individual level + Heterogeneity (MTE)

⬇

4. (Mis)allocation

Within department variation

$$Y_i = \alpha_0 + \alpha_1 \text{Industry}_i + \delta_{mu} + \alpha_2 \mathbf{X_i} + \varepsilon_i \tag{1}$$

- δ_{mu}: PhD major $\times$ university FE

Within department variation

$$Y_i = \alpha_0 + \alpha_1 \text{Industry}_i + \delta_{mu} + \alpha_2 \mathbf{X_i} + \varepsilon_i \quad (1)$$

- δ_{mu}: PhD major $\times$ university FE

	Log(1+Publications (stock))				
	(1)	(2)	(3)	(4)	(5)
Industry	-0.911***	-0.943***	-0.936***	-0.942***	-0.900***
	(0.0200)	(0.0255)	(0.0255)	(0.0278)	(0.0296)
Observations	21,406	21,406	21,406	21,406	21,406
Demographics	Yes	Yes	Yes	Yes	Yes
Experience	Yes	Yes	Yes	Yes	Yes
PhD major FE		Yes	Yes	Yes	Yes
Doct. Inst. FE			Yes	Yes	Yes
PhD major x Doct. Inst. FE				Yes	Yes
Taste					Yes

Standard errors in parentheses clustered at the doctoral institution level

Within department variation

$$Y_i = \alpha_0 + \alpha_1 \text{Industry}_i + \delta_{mu} + \alpha_2 \mathbf{X_i} + \varepsilon_i \tag{2}$$

- δ_{mu}: PhD major $\times$ university FE

	Log(1+Publications (stock))				
	(1)	(2)	(3)	(4)	(5)
Industry	-0.911***	-0.943***	-0.936***	-0.942***	-0.900***
	(0.0200)	(0.0255)	(0.0255)	(0.0278)	(0.0296)
Observations	21,406	21,406	21,406	21,406	21,406
	Log(1+Earnings)				
	(1)	(2)	(3)	(4)	(5)
Industry	0.367***	0.296***	0.295***	0.302***	0.302***
	(0.0121)	(0.0134)	(0.0135)	(0.0149)	(0.0169)
Observations	21,406	21,406	21,406	21,406	21,406
Demographics	Yes	Yes	Yes	Yes	Yes
Experience	Yes	Yes	Yes	Yes	Yes
PhD major FE		Yes	Yes	Yes	Yes
Doct. Inst. FE			Yes	Yes	Yes
PhD major x Doct. Inst. FE				Yes	Yes
Taste					Yes

Standard errors in parentheses clustered at the doctoral institution level

Overview

1. ‘Control’ for ability: within department variation

⬇

2. Selection at the individual level (LATE)

⬇

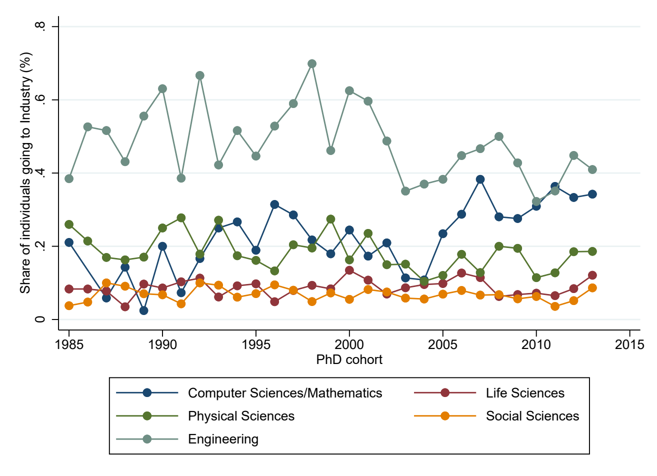
3. Selection at the individual level + Heterogeneity (MTE)

⬇

4. (Mis)allocation

Instrumental Variable

- **Ideal Variation**: PhDs graduating face an exogenous demand from firms' and an exogenous demand from universities'
- **Empirical Variation**: exploits inter-temporal variation in firms' vs universities' demand for PhDs in the same major (within major across cohorts)



Instrumental Variable

- **Instrument**: exploits inter-temporal variation in firms' vs universities' demand for PhDs in the same major (within major across cohorts)
- **Empirically**: PhD major $\times$ PhD cohort FE, controlling for PhD major FE
  - I create a leave-one out continuous measure $Z_i =$ share of individuals in major/cohort going to Ind.

Instrumental Variable

- **Instrument**: exploits inter-temporal variation in firms' vs universities' demand for PhDs in the same major (within major across cohorts)
- **Empirically**: PhD major $\times$ PhD cohort FE, controlling for PhD major FE
    - I create a leave-one out continuous measure $Z_i =$ share of individuals in major/cohort going to Ind.
- **Identifying assumption**: individuals that are more/less productive in Ind. (vs Acad.) do not time when they go on the market based on macro. conditions
- **Controls**:
    - Doctoral institution FE
    - Indicator for having published prior to graduation

Instrumental Variable

- **Instrument**: exploits inter-temporal variation in firms' vs universities' demand for PhDs in the same major (within major across cohorts)
- **Empirically**: PhD major $\times$ PhD cohort FE, controlling for PhD major FE
  - I create a leave-one out continuous measure Z_i = share of individuals in major/cohort going to Ind.
- **Identifying assumption**: individuals that are more/less productive in Ind. (vs Acad.) do not time when they go on the market based on macro. conditions
- **Controls**:
  - Doctoral institution FE
  - Indicator for having published prior to graduation
- **Test**: conditions at the time of PhD entry *do not* predict sorting

Instrumental Variable - First-stage

	Industry		
	(1)	(2)	(3)
Z_i	0.220***	0.214***	0.214***
	(0.0385)	(0.0390)	(0.0389)
Demographics	Yes	Yes	Yes
Experience	Yes	Yes	Yes
PhD major FE	Yes	Yes	Yes
Doct. inst. FE		Yes	Yes
Publ. bef. grad.			Yes
F-stat	32.5	30.1	30.1
Observations	21,406	21,406	21,406
R-sq	0.163	0.185	0.185

Notes: Standard errors (in parentheses) are two-way clustered at the PhD major and PhD cohort level.

- Z_i = share of individuals in major/cohort going to Ind.
- If everyone in her major/cohort goes to Industry, the focal individual has a 21% higher probability of going to Ind. (vs Acad.)

Instrumental Variable - 2SLS results

	Log(1+Publications)		Log(1+Earnings)	
	OLS	2SLS	OLS	2SLS
Industry	-0.900***	-1.227***	0.274***	0.527**
	(0.0296)	(0.476)	(0.0147)	(0.313)
F-stat		30.1		30.1
Observations		21,406		21,406
Demographics		Yes		Yes
Experience		Yes		Yes
PhD major FE		Yes		Yes
Doct. Inst. FE		Yes		Yes
Publ. bef. grad.		Yes		Yes

Notes: Standard errors in parentheses clustered at the PhD major and PhD cohort level * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

- OLS $\neq$ 2SLS because (Ishimaru, 2024):
  - Covariate weight difference
  - Treatment-level weight difference
  - **Selection**

Overview

1. 'Control' for ability: within department variation

⬇

2. Selection at the individual level (LATE)

⬇

3. Selection at the individual level + Heterogeneity (MTE)

⬇

4. (Mis)allocation

Heterogeneity on observables and unobservables - MTE

- I estimate **Marginal Treatment Effects** (MTE)
  - The ATE for groups of individuals who have the same **latent propensity** to be treated based on observables and unobservables
  - Is the micro-component of the ATE, ATT, LATE
- MTE for individuals with observables $X = x$ and unobservable $U_D = u$ is:

$$\mathrm{MTE}(X = x, U_D = u) = \mathbb{E}[Y^I - Y^A | X = x, U_D = u] \tag{3}$$

Heterogeneity on observables and unobservables - MTE

- I estimate **Marginal Treatment Effects** (MTE)
  - The ATE for groups of individuals who have the same **latent propensity** to be treated based on observables and unobservables
  - Is the micro-component of the ATE, ATT, LATE
- MTE for individuals with observables $X = x$ and unobservable $U_D = u$ is:

$$\mathrm{MTE}(X = x, U_D = u) = \mathbb{E}[Y^I - Y^A | X = x, U_D = u] \tag{3}$$

- I can predict expected treatment effects for every individual based on their observables, treatment status and unobservable

Individual-specific expected treatment effects

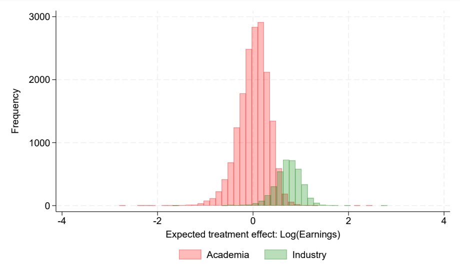


Figure 2: Log(Earnings)

- Individuals who start their career in Industry have higher earnings gains of joining Ind. (vs Acad.)

Individual-specific expected treatment effects

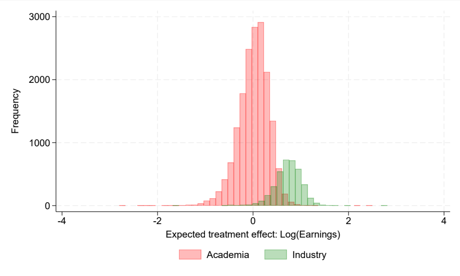


Figure 2: Log(Earnings)

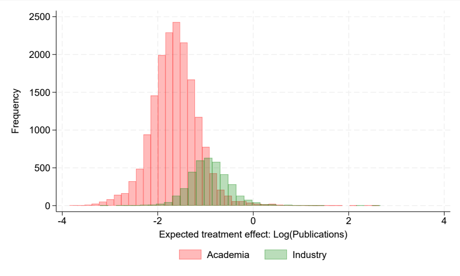


Figure 3: Log(Publications)

- Individuals who start their career in Industry have higher earnings gains of joining Ind. (vs Acad.)

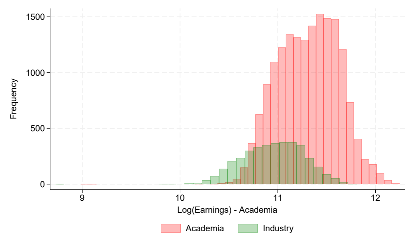
- Individuals who start their career in Industry have a lower publication decrease of joining Ind. (vs Acad.)

MTE Propensity Score

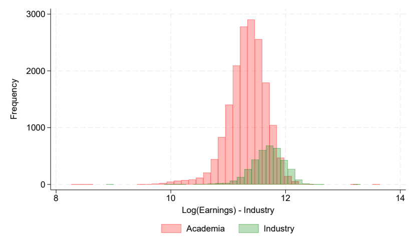
MTE curves

Individual-specific expected potential outcomes

Earnings in Academia



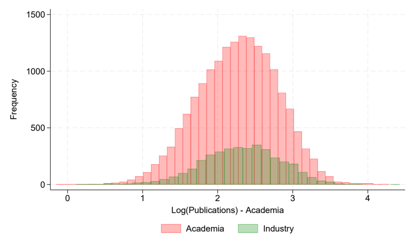
Earnings in Industry



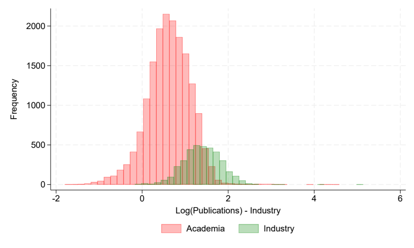
- Individuals who start their career in Industry have higher earnings in Industry and lower earnings in Academia than those who start their career in Academia

Individual-specific expected potential outcomes

Publications in Academia



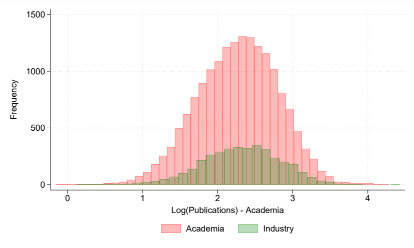
Publications in Industry



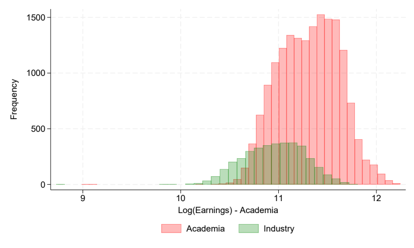
- Individuals who start their career in Industry have a higher publication potential in Industry than those who start their career in Academia
- In the private sector, individuals who start their career in Industry join bigger firms than those who start their career in Academia

Individual-specific expected potential outcomes

Publications in Academia



Earnings in Academia



- Individuals who start their career in Industry have the same publication potential in Academia than those who start their career in Academia
- Individuals who start their career in Industry lack non-research related skills that are valued in Academia

Overview

1. 'Control' for ability: within department variation $\rightarrow$ ATE

⬇

2. Selection at the individual level $\rightarrow$ LATE

⬇

3. Selection at the individual level + Heterogeneity $\rightarrow$ MTE

⬇

4. (Mis)allocation

(Mis)Allocation

- Goal: think about the **wedge** between private and social returns
  - Individuals' sorting is influenced by their private returns in Ind. (vs Acad.)
  - Individuals' 'best use' is influenced by their social returns in Ind. (vs Acad.)
  - The wedge between private and social returns comes from externalities

(Mis)Allocation

- Goal: think about the **wedge** between private and social returns
  - Individuals' sorting is influenced by their private returns in Ind. (vs Acad.)
  - Individuals' 'best use' is influenced by their social returns in Ind. (vs Acad.)
  - The wedge between private and social returns comes from externalities
- Goal: highlight **trends** and **patterns**
  - I cannot capture the entirety of private returns
  - I cannot capture the entirety of externalities
  - I can only discuss the *rate* of externalities, not the *direction*

(Mis)Allocation

- Goal: think about the **wedge** between private and social returns
    - Individuals' sorting is influenced by their private returns in Ind. (vs Acad.)
    - Individuals' 'best use' is influenced by their social returns in Ind. (vs Acad.)
    - The wedge between private and social returns comes from externalities
- Goal: highlight **trends** and **patterns**
    - I cannot capture the entirety of private returns
    - I cannot capture the entirety of externalities
    - I can only discuss the *rate* of externalities, not the *direction*
- I am currently trying to build a simple model of **social returns** where I use the individual-level MTE for earnings and publications to shed light on misallocation
    - I can compare the **wedge in earnings** with the **wedge in publications**
    - I can **characterize** individuals based on where they are in the Earnings and Publications distribution in Industry and in Academia
- **All the extensions are the fun part!**

Conclusion

In this paper, I study the **performance** of scientists in Industry and Academia after graduation:

- **Data**: I use administrative surveys that give me individual-level information about two measures of performance: **earnings** and **publications**
- **Identification**:
  - I account for individual-level **selection** with an instrument that leverage variation in firms' vs universities' demand over time for PhDs in the same major
  - I incorporate **heterogeneity** with the estimation of MTE
- **Next steps**: use the reduced-form individual-level estimates to highlight trends and patterns in the misallocation of scientists across sectors

Thank you!

MTE - Parameters

$$ATE(x) = \int_0^1 MTE(x,u)du \tag{4}$$

$$ATT(x) = \int_0^1 h_{ATT}(x,u)MTE(x,u)du \tag{5}$$

$$ATUT(x) = \int_0^1 h_{ATUT}(x,u)MTE(x,u)du \tag{6}$$

$$IV(x) = \int_0^1 h_{IV}(x,u)MTE(x,u)du \tag{7}$$

back

MTE Propensity score

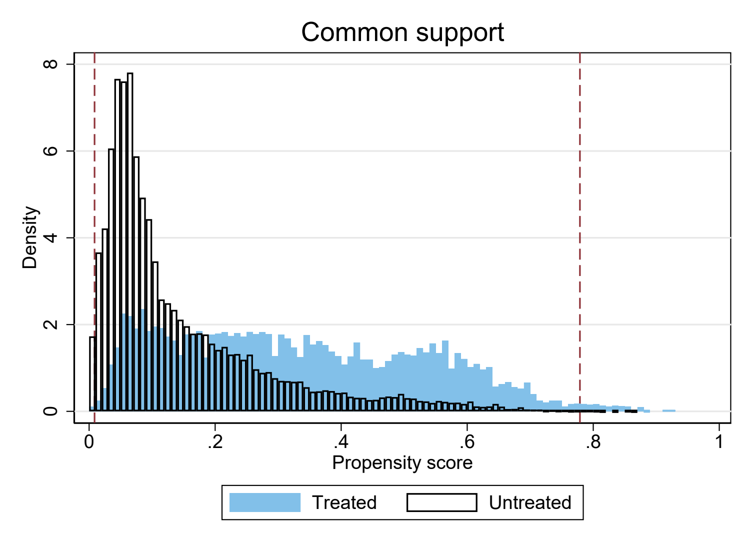


Figure 4: Propensity Score

MTE Curves

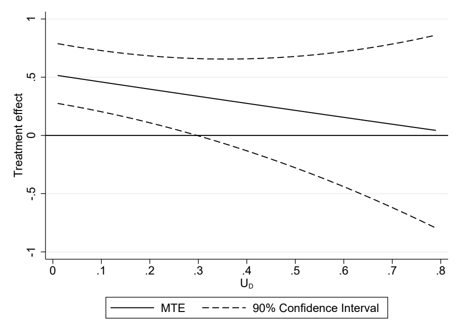


Figure 5: Log(1+Earnings)

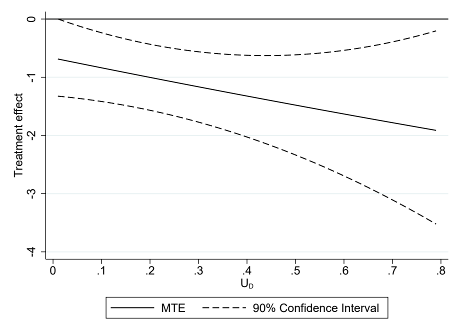


Figure 6: Log(1+Publications)

back

References

Rajshree Agarwal and Atsushi Ohyama. Industry or academia, basic or applied? career choices and earnings trajectories of scientists. *Management Science*, 59 (4):950–970, 2013.

Philippe Aghion, Mathias Dewatripont, and Jeremy C Stein. Academic freedom, private-sector focus, and the process of innovation. *The RAND Journal of Economics*, 39(3):617–635, 2008.

Ufuk Akcigit and Nathan Goldschlag. Where have all the" creative talents" gone? employment dynamics of us inventors. Technical report, National Bureau of Economic Research, 2023.

Ufuk Akcigit, Jeremy G Pearce, and Marta Prato. Tapping into talent: Coupling education and innovation policies for economic growth. Technical report, National Bureau of Economic Research, 2020.

Alex Bell, Raj Chetty, Xavier Jaravel, NEVIANA Petkova, and John Van Reenen. Who becomes an inventor in america. *The Importance of exposure to innovation*, 66, 2017.

Murat Alp Celik. Does the cream always rise to the top? the misallocation of talent in innovation. *Journal of Monetary Economics*, 133:105–128, 2023.

Russell Coff and David Kryscynski. Invited editorial: Drilling for micro-foundations of human capital–based competitive advantages. *Journal of management*, 37(5):1429–1443, 2011.

Partha Dasgupta and Paul A David. Toward a new economics of science. *Research policy*, 23(5):487–521, 1994.

Jeffrey L Furman and Scott Stern. Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review*, 101 (5):1933–1963, 2011.

Shulamit Kahn and Donna K Ginther. The impact of postdoctoral training on early careers in biomedicine. *Nature biotechnology*, 35(1):90–94, 2017.

Michael Roach and Henry Sauermann. A taste for science? phd scientists' academic orientation and self-selection into research careers in industry. *Research policy*, 39(3):422–434, 2010.

Michael Roach and Henry Sauermann. Can technology startups hire talented early employees? ability, preferences, and employee first job choice. *Management Science*, 2023.

Michael Roach and Henry Sauermann. Ability, preferences, and stem doctorate early career choices. *Working Paper*, 2024.

Paul M Romer. Endogenous technological change. *Journal of political Economy*, 98(5, Part 2):S71–S102, 1990.

Henry Sauermann and Michael Roach. Science phd career preferences: levels, changes, and advisor encouragement. *PloS one*, 7(5):e36307, 2012.

Henry Sauermann and Paula Stephan. Conflicting logics? a multidimensional view of industrial and academic science. *Organization science*, 24(3):889–909, 2013.

Pian Shu. Innovating in science and engineering or'cashing in'on wall street? evidence on elite stem talent. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (16-067), 2016.

Paula Stephan. *How economics shapes science*. Harvard University Press, 2012.

Paula E Stephan. The economics of science. *Journal of Economic literature*, 34(3):1199–1235, 1996.

Paula E Stephan, Albert J Sumell, Grant C Black, and James D Adams. Doctoral education and economic development: The flow of new ph. ds to industry. *Economic Development Quarterly*, 18(2):151–167, 2004.

Scott Stern. Do scientists pay to be scientists? *Management science*, 50(6):835–853, 2004.