# Producing Science

Julia Lane, Jacques Mairesse, Michele Pezzoni and Paula Stephan

Alfred P. Sloan Foundation, July 24, 2013

# Key ideas

- Current state of economics of science
  - One off studies
  - One dimensional
  - Limited unit of analysis
- Caltech project
  - Builds on new data at project level
  - Allows us to model productivity
- National and international activities
  - CIC
  - International

# Outline

- Current State: Economics of Science
  - What we know
  - How we know it
  - What we don't know
  - The potential
- Caltech Project
  - Framework
  - Data Description
  - Empirical Analysis
- National and International Activities
- Next Steps

# Outline

- Current State: Economics of Science
  - What we know
  - How we know it
  - What we don’t know
  - The potential
- Caltech Project
  - Framework
  - Data Description
  - Empirical Analysis
- National and International Activities
- Next Steps

# What We Know

- Knowledge produced by university researchers embodied in multiple forms:
  - Publications
  - Patents
  - Training and placement of students

# What We Know…

- University knowledge often produced in labs that resemble firms; directed and overseen by PI’s
  - Initially get start-up package; Hire and fire and select skill-mix of individuals; Raise funds to keep lab functioning
- Majority of funds for lab come from federal government
- Knowledge production involves multiple inputs: Q=f(k, t, m, e, s)
  - Some inputs, such as knowledge and time, are embodied in people

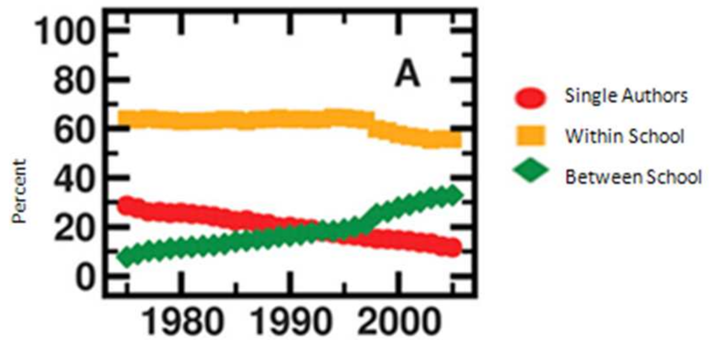
# What We Know…

- Characteristics of productive people in terms of location, age, gender, country of birth
- Certain characteristics of productive teams
- Size of teams as measured by metrics such as co-inventors and coauthors over time
- Composition of team in terms of position
- Placement of new PhDs at firms
- Speed with which knowledge leaks out

# What we know: Size of Teams

- Size of team growing over time
- Multi-University Collaboration increasing
- Within school declining
- Sole-authored articles are rapidly becoming dianosaurs
- Work by Jones et. Al.

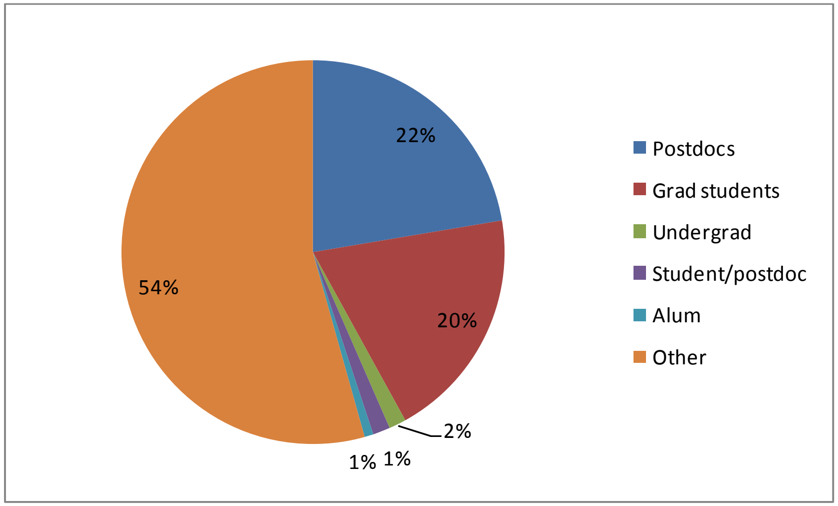
The Rise in Multi-University Collaboration, Science and Engineering

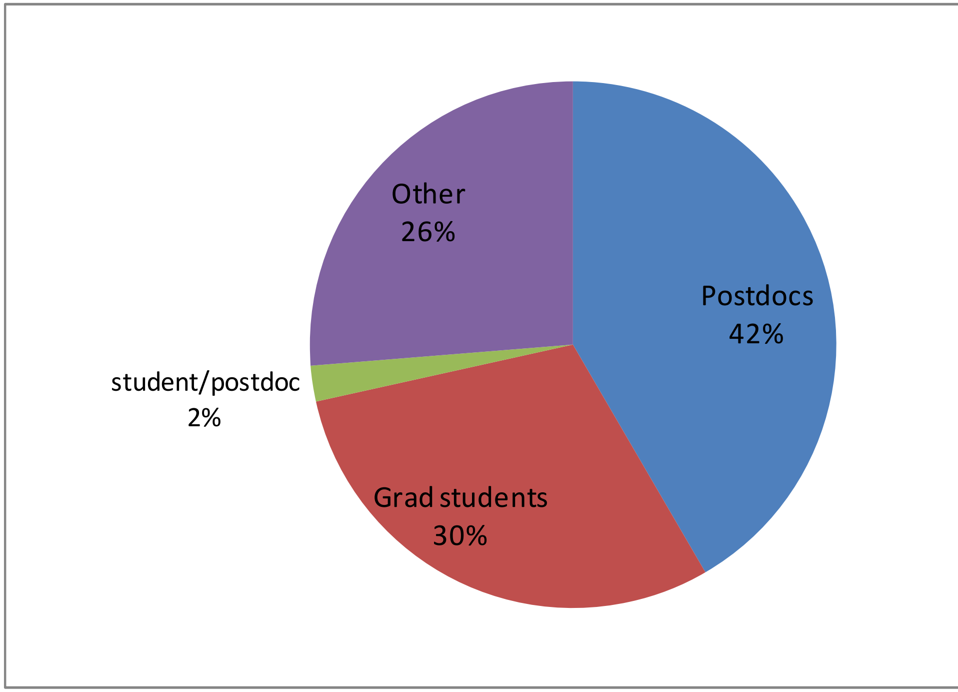



Source: B. F. Jones et al., Science 322, 1259 -1262 (2008)

# What we know: Composition of Team

- Study of authorship patterns in journal *Science* 2007

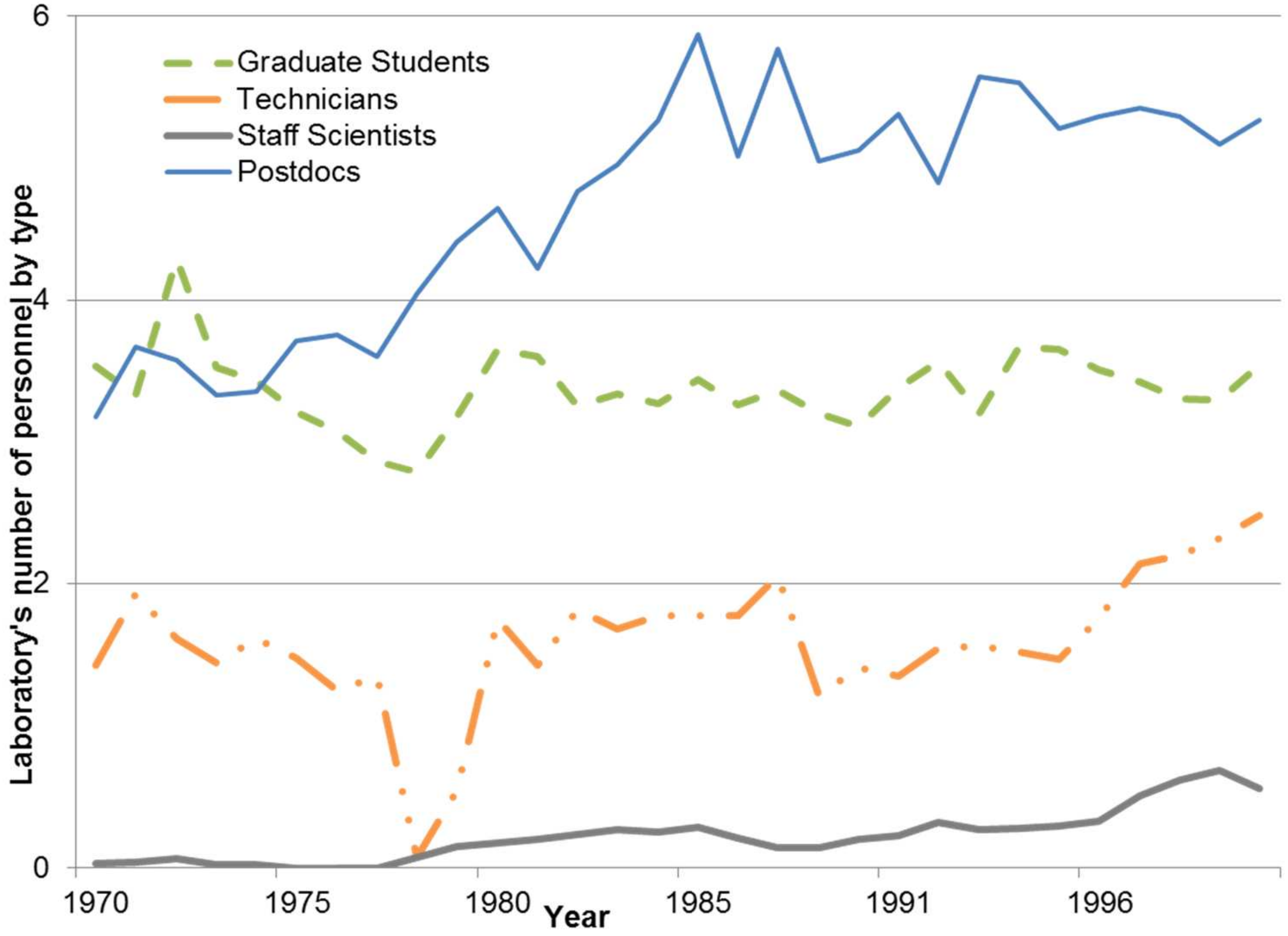



First Authors

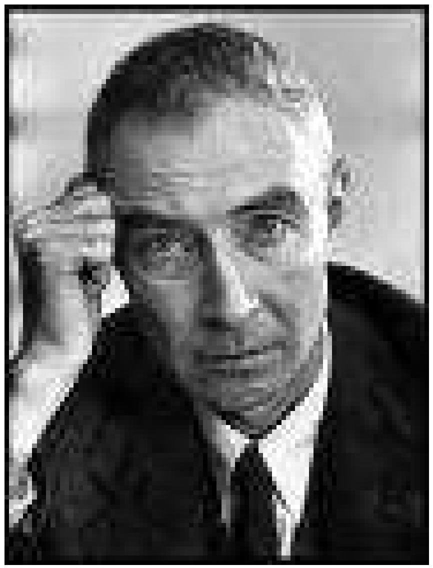
All Authors

# What we know:Lab Composition Over Time MIT Dept. of Biology




Conti and Liu

# What we know: Placement of PhD Students Going to Industry



"The best way to send information is to wrap it up in a person"
*J. Robert Oppenheimer*

- Top 32 hiring firms by NAICS classification
- Greatest number of hires were in firms working in computer and electrical
- Followed by hires in publishing industries and professional scientific and technical services

# What we know: Role of Small Firms

- Placement findings suggest that small firms play a larger role in innovation than R&D data would suggest
  - Top 200 R&D firms expend more than 70% of all R&D in U.S.
  - Hire only 40% of new PhDs

# What we know: Role of Specialization

- Annamaria Conti and Fabiana Visentin examine relationship of specialization to publication output for labs at EPFL
- Measure specialization by topic-modeling dissertation abstracts of students in lab
- Find productivity of lab relates to specialization of team
- Relationship is inverted “U;” peaks at 52 unique topics

# Outline

- Current State: Economics of Science
  - What we know
  - How we know it
  - What we don’t know
  - The potential
- Caltech Project
  - Framework
  - Data Description
  - Empirical Analysis
- National and International Activities
- Next Steps

# How we know it: Sources

- Federal surveys (SED, SDR); publication and patent data bases
- Observation
- Interviews
- Paper trail—citations to patents and publications; co-authorship patterns on the two
- Surveys
- Linking data sets to one another, such as publications and citations
- Special data requests

# How we know it: One off studies

- Freeman *et al.* study of changing patterns of international collaboration in research (Survey, corresponding authors, 3 time periods)
- Conti and Liu study of MIT labs; Conti study of EPFL labs (administrative records; LinkedIn data)
- Sauermann and Roach study of work aspirations of graduate students and postdocs (survey)
- Tambi study of IT skills using “digital breadcrumbs”
- Thursby and Thursby study of ownership of patents invented by university faculty (Match 1995 NRC faculty data with patent data)

# How we know it: One off studies

- Furman and Stern study of how deposit of research materials at BRCs affects use of materials and diffusion of materials (link deposits to articles and citations)
- Ding *et al.* study of relationship of gender to patenting, SAB membership, etc. (Proquest Dissertation data linked with publicly available data, including patents)
- Murray *et al*. study of how intellectual property rights affect diffusion and use of mouse models (patent data; citations to relevant "mouse" articles; nearest neighbor citations)
- Own-Smith study of relationship of physical proximity to productivity (floor plans and publication data)

# How we know it: One off studies

- Levin and Stephan study of relationship of age, period and cohort effects to productivity
- Uzzi and Jones study of authorship patterns over time
- Franzoni *et al.* study of how changing incentives relate to submission patterns to *Science*
- Stephan *et al.* study of firm placement of new PhD

# What we know:
# Problems with One-off Approach

- Herculean effort
- Costly
- Non-reproducible
- Quickly becomes out of date
- Confidentiality issues restrict use
- Difficult to reproduce

# Example: Case Study

- 1980 Levin and Stephan proposed using SDR to study relationship between productivity to age, time, and cohort
  - "Is Science a Young Person's Game?"
  - Obvious database was SDR
  - Lacked publication data
  - 1983 received funding to have NRC match SDR for 1973-1979 to ISI data (Thank you, Sloan!)
  - Costly; slow; non-sharable
  - Research published in 1992—12 years after initial request

# Outline

- Current State: Economics of Science
  - What we know
  - How we know it
  - What we don’t know
  - The potential
- Caltech Project
  - Framework
  - Data Description
  - Empirical Analysis
- National and International Activities
- Next Steps

# What we don't know: Examples

- Current analysis often one-dimensional
  - Relationship of productivity to team characteristics
  - To proximity
  - To age
  - To gender
- But production of research is not one dimensional
- Involves multiple inputs: $Q=f(k, t, m, e, s)$
- Know there are other inputs, but focus is almost exclusively on the "t" and the "k" which are embodied in people
- With but rare exception, ignore equipment, materials and space; characteristics of team members
- Ignore prices of inputs, be they wages or costs of other inputs—yet clear that costs affect hiring decisions

# What we don’t know

- “One dimensional focus” means we do not apply what we have learned from productivity studies of firms to productivity of labs
- Yet labs are much like firms
- PIs much like entrepreneurs
- Number of questions that we need to address

# Outline

- Current State: Economics of Science
  - What we know
  - How we know it
  - What we don't know
  - The potential
- Caltech Project
  - Framework
  - Data Description
  - Empirical Analysis
- National and International Activities
- Next Steps

# The potential: Star Metrics Capability

- Position of individuals working on project; team (faculty, staff scientist, postdoc, graduate student)
- Characteristics of individuals working on the project (year of degree, former institutions, gender; expertise from CV data, digital bread crumbs)
- Publications; patents, citations to those
- Topics of research (topic modeling of grants, publications, dissertations, and patents)
- Placement of PhDs and postdocs (LEHD data; LinkedIn)

# Could SM Data Have Been Used?

- Freeman et al. study of international collaboration in research (Survey)
- Conti and Liu study of MIT labs; Conti and Visentin study of EPFL labs (Administrative records; LinkedIn, topic modeling dissertation abstracts)
- Sauermann and Roach study of work aspirations of graduate students and postdocs (survey)
- Thursby and Thursby study of ownership of patents invented by university faculty (NRC 1995 data of faculty names matched with patent data)
- Tambi approach of using LinkedIn data to measure job outcomes and job skills
- Furman and Stern study of how deposit of research materials at BRCs affects use of materials and diffusion of materials (link deposits to articles and citations)
- Ding et al. study of relationship of gender to patenting, SAB membership, etc. (Proquest Dissertation data linked with publicly available data, including patents)
- Murray et al. study of how intellectual property rights affect diffusion and use of mouse models (citations to relevant “mouse” articles; nearest neighbor citations)
- Own-Smith study of relationship of physical proximity to productivity (floor plans and publication data). Remember: Star Metrics is a platform—we can hang anything on it that universities make available
- Stephan and Levin study of age, cohort, time period and productivity (SDR matched with ISI data)
- Ginther studies of how publication and promotion relate to gender
- Stephan placement of new PhDs (Verbatim records from SED of first placement; Star Metrics linked with LEHD, other sources will provide longituninal picture)

# The potential

- Star Metrics would dramatically cut down on number of “one-off” studies
- Increase ability to replicate and build on other’s research
- Increase timeliness of research

# Outline

- Current State: Economics of Science
  - What we know
  - How we know it
  - What we don’t know
  - The potential
- Caltech Project
  - Framework
  - Data Description
  - Empirical Analysis
- National and International Activities
- Next Steps

# The Institution

- Size
  - Approximately 300 faculty
  - 980 undergraduates
  - 1250 graduate students
  - 650 (approximate) postdocs
- Focus
  - Science and engineering
  - Minimal focus on humanities and social sciences

# Outline

- Current State: Economics of Science
  - What we know
  - How we know it
  - What we don't know
  - The potential
- Caltech Project
  - Framework
  - Data Description
  - Empirical Analysis
- National and International Activities
- Next Steps

# The Empirical Framework

- Multi dimensional production of research
- Involves multiple inputs: $Q=f(k, t, m, e, s)$
- Include equipment, materials and space; characteristics of team members
- Include prices of inputs, be they wages or costs of other inputs—since costs affect hiring decisions
- Include disciplinary information

# The Qualitative Framework

- Three days at Caltech in early January
  - Met with Mory Gharib, Hans W. Liepmann Professor of Aeronautics & Professor of Bioinspired Engineering, Vice Provost for Research
  - Dick Seligman, Associate Vice President for Research Administration
  - David Mayo, Director, Office of Sponsored Research
- Five days at Caltech in early April
  - Conducted in depth interviews with 6 faculty
  - Met with Caltech team

# Outline

- Current State: Economics of Science
  - What we know
  - How we know it
  - What we don't know
  - The potential
- Caltech Project
  - Framework
  - Data Description
  - Empirical Analysis
- National and International Activities
- Next Steps

# Data

- Administrative records of Federal grants for period 2000-2012
- Files
  - Employee file
  - Vendor file
  - Subcontract file
  - Overhead file

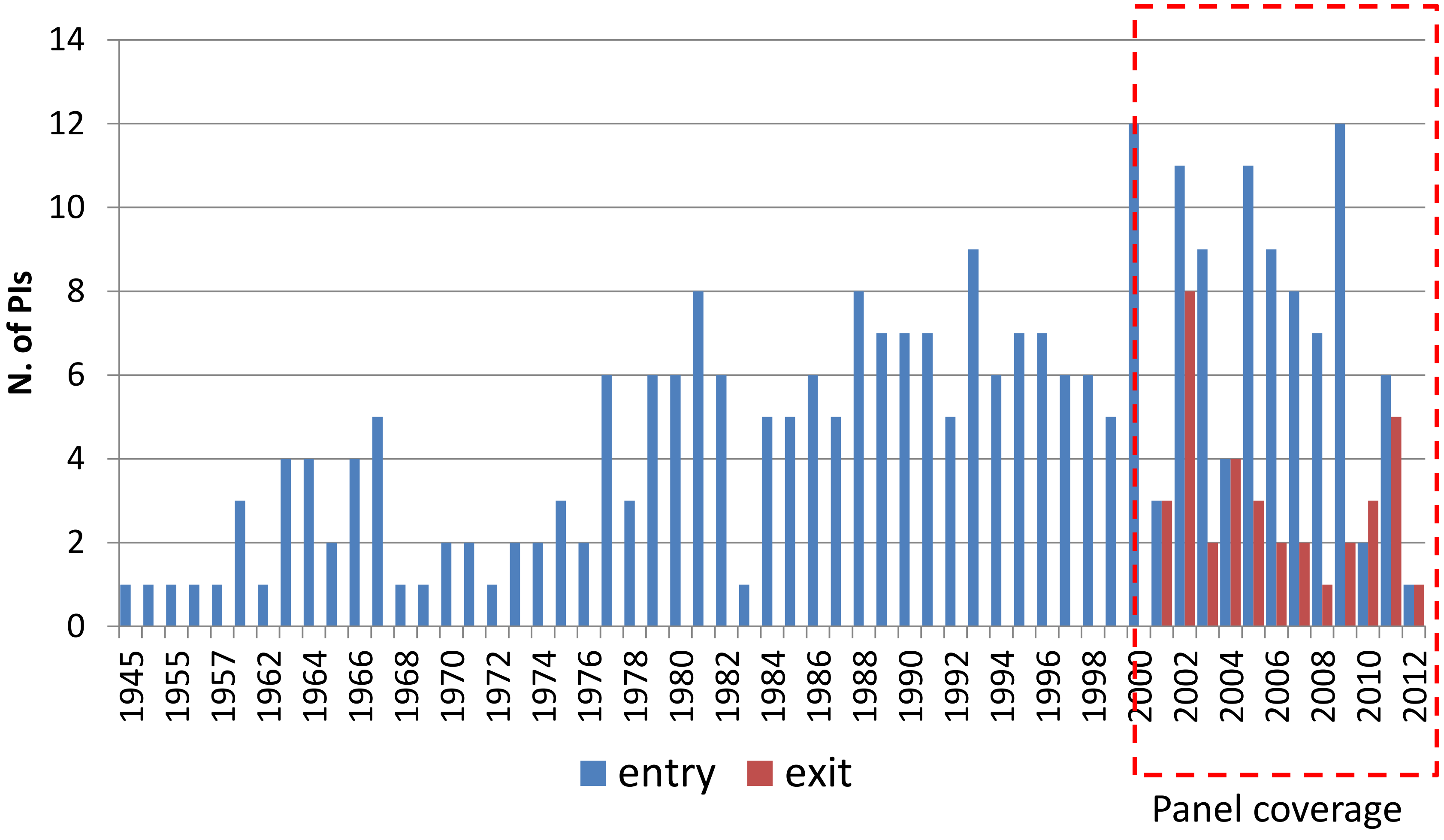
# Matched Data

- Publications and citations
- Patents and citations
- Dissertations
- Topic modeling of grants, dissertations and articles
- Placement data (eventually) from matching dissertation awards to LEHD

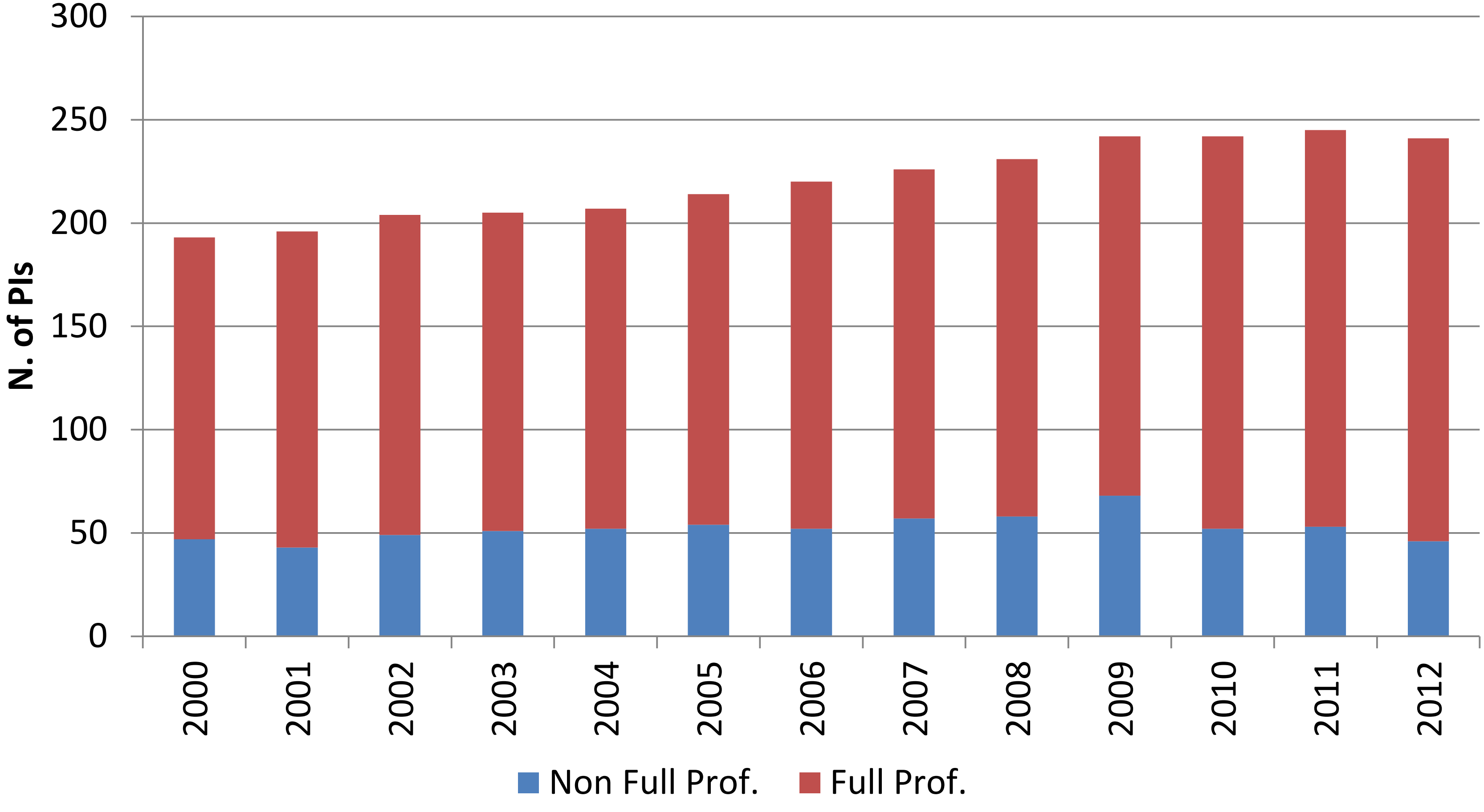
# Panel Data description

- 276 PIs
  - Limited to faculties for which we have CV information
  - Active during 2000-2013 (PIs retiring/leaving Caltech before 2000 are not considered)
- 60% of the PIs are active from 2000 to 2013 (The panel covers on average 11.25 years per PI)
- We present statistics on: entry&exit, career, research group composition, productivity, funding and expenses, labor&capital, and PI’s personal characteristics

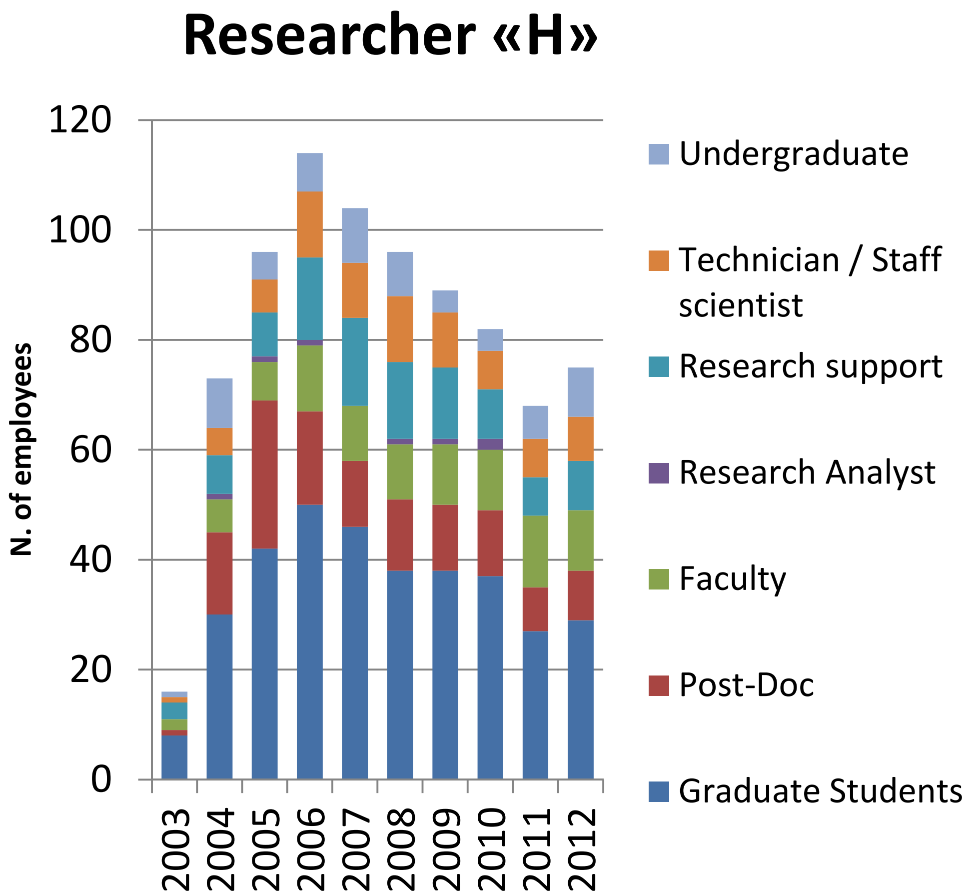
# The Data Coverage

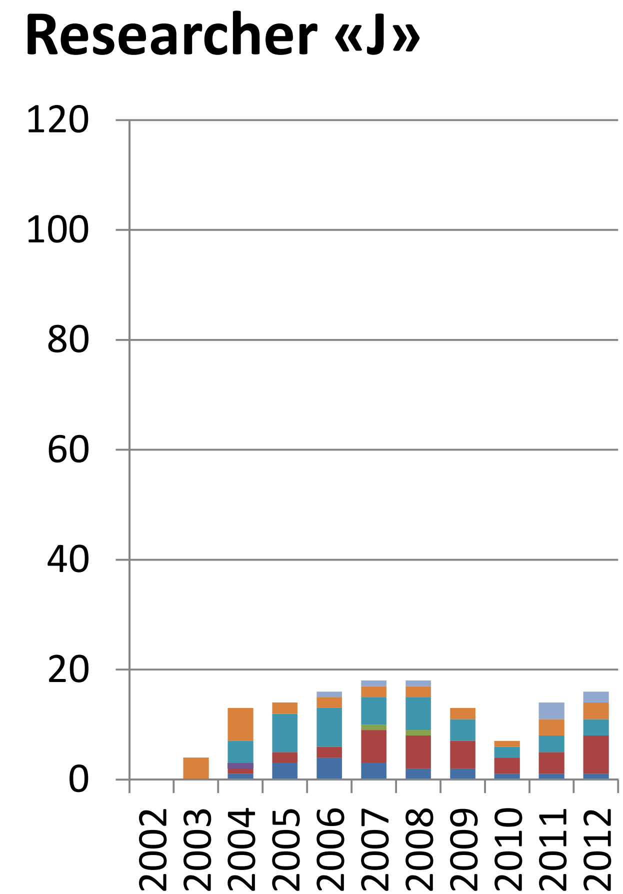


# The Sample Size

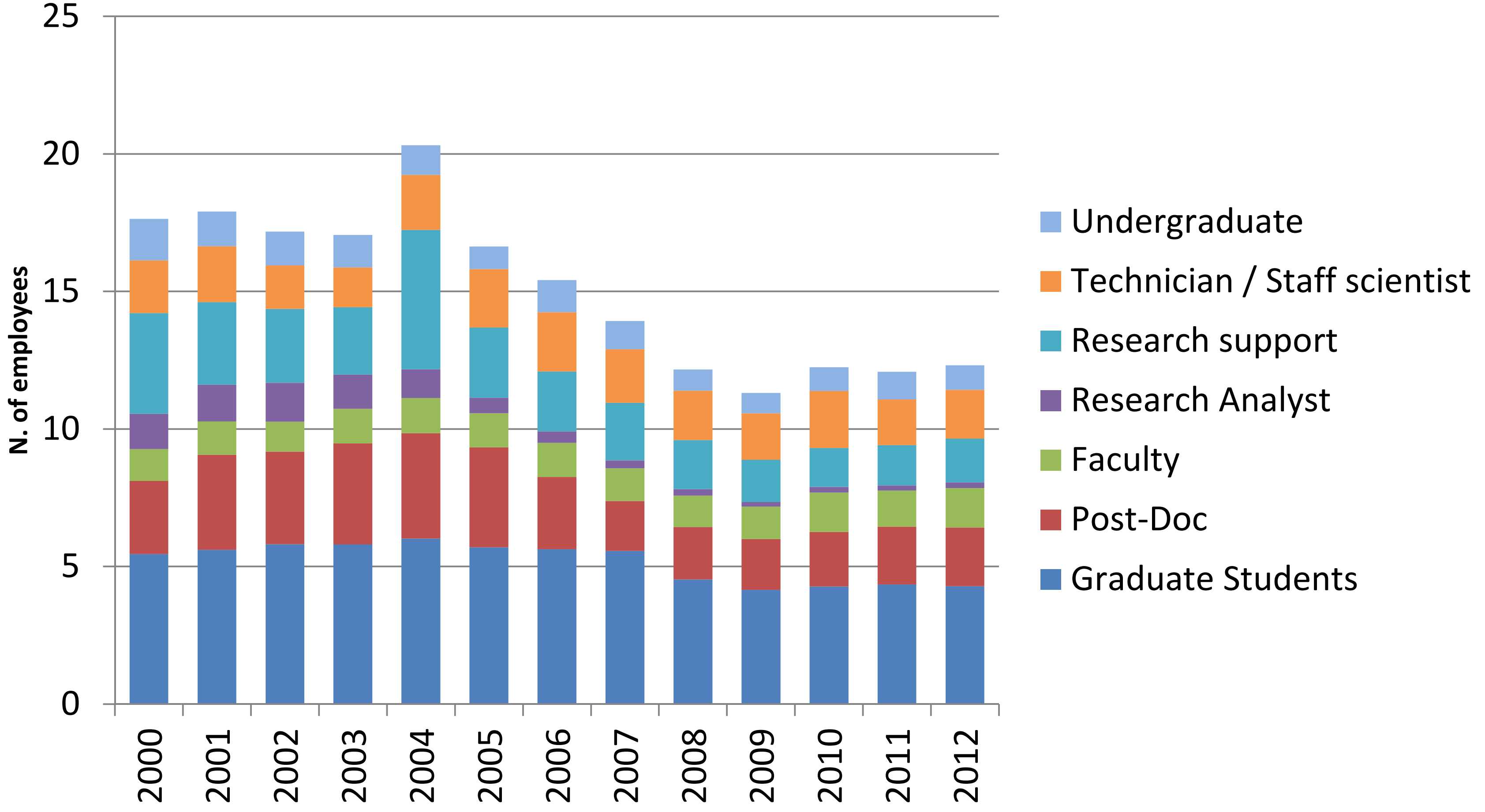


# Research teams: Varies substantially in size and composition

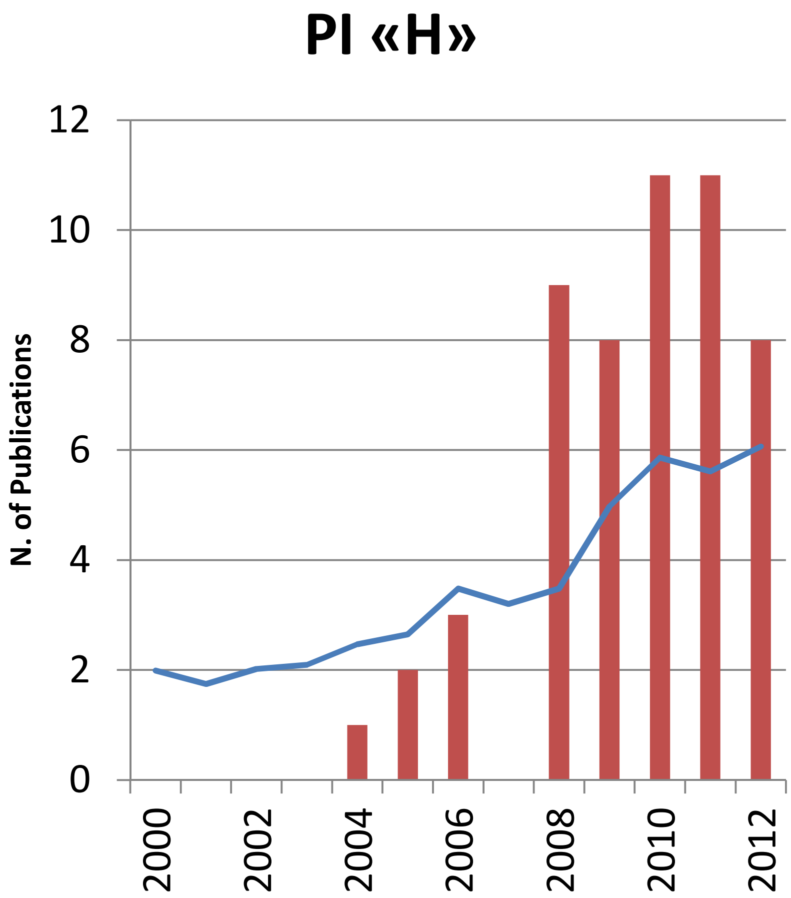


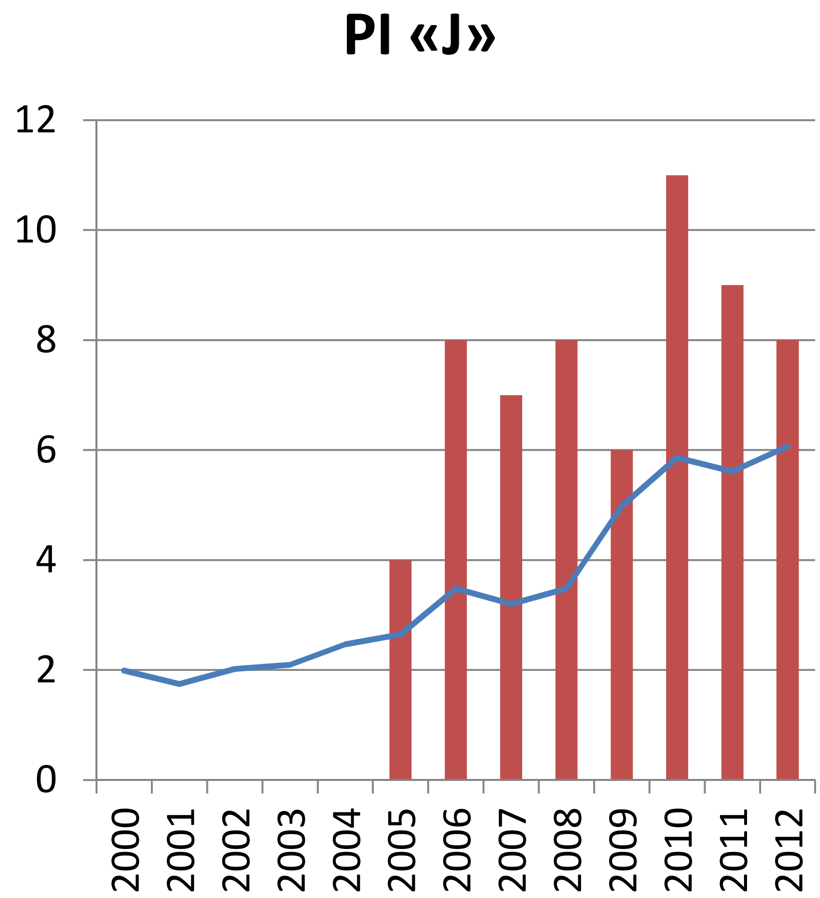



# Research group composition:
## average per PI



# Number of Publications



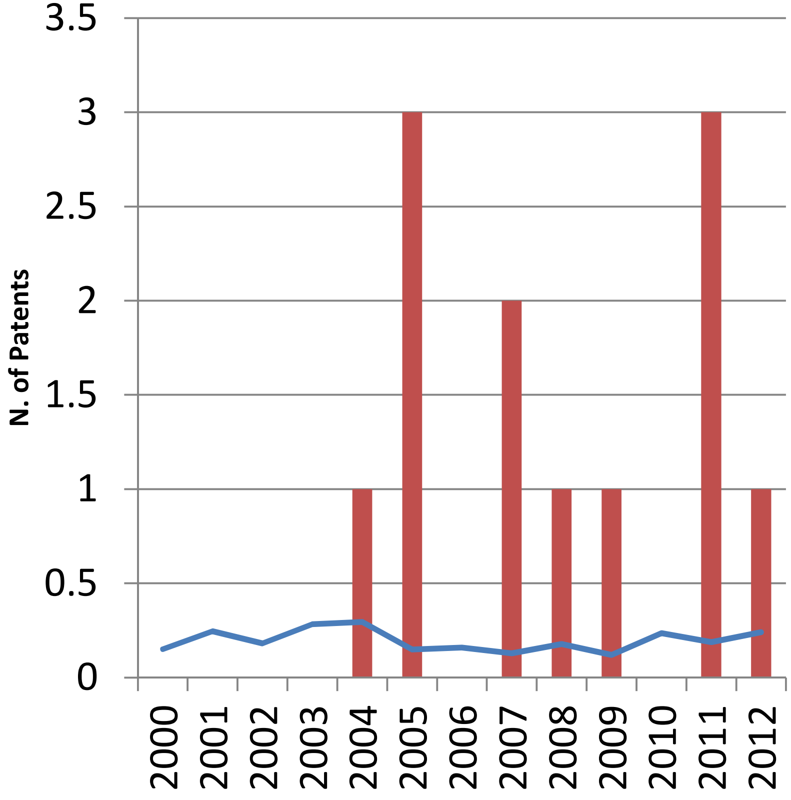




*blue solid line is the average per PI*

# Number of Patents

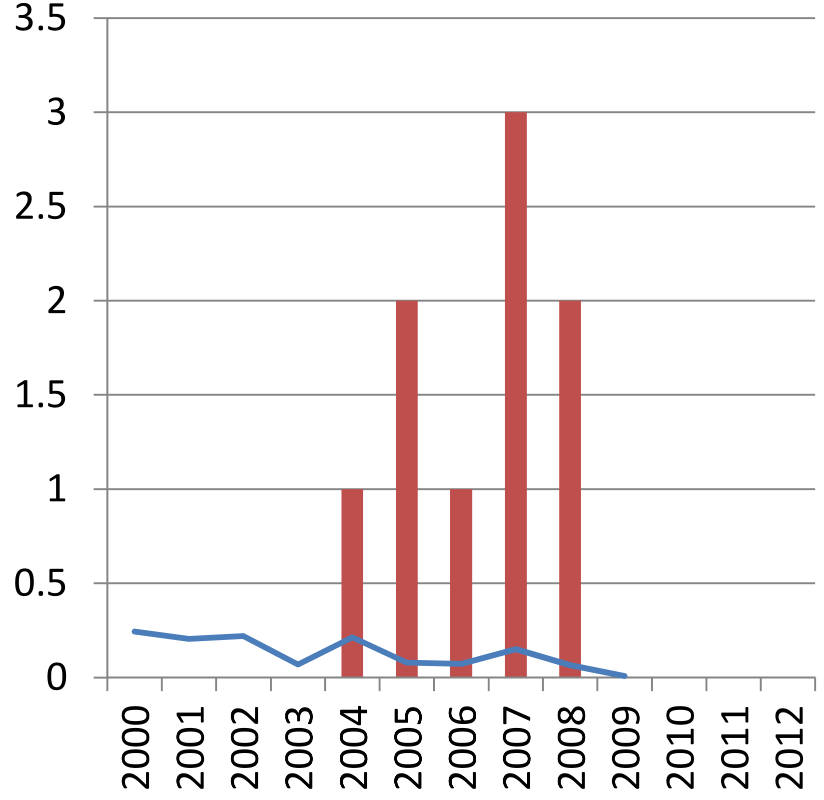
*PI «J» does not have patents*

## PI «H» (USPTO patents*)



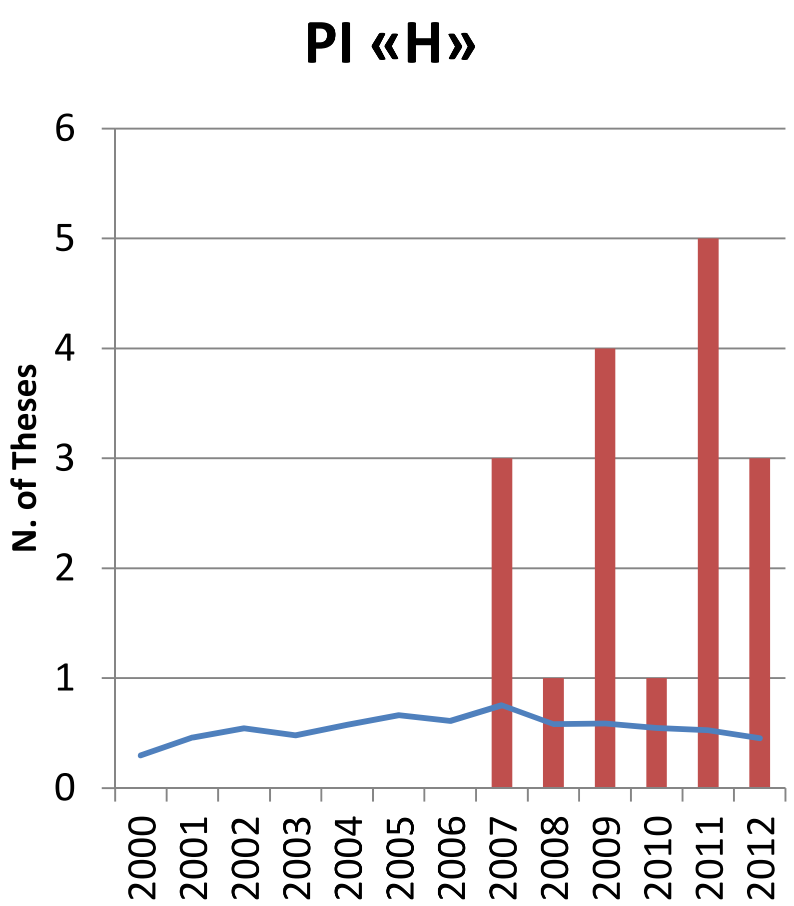

* patents by granting date

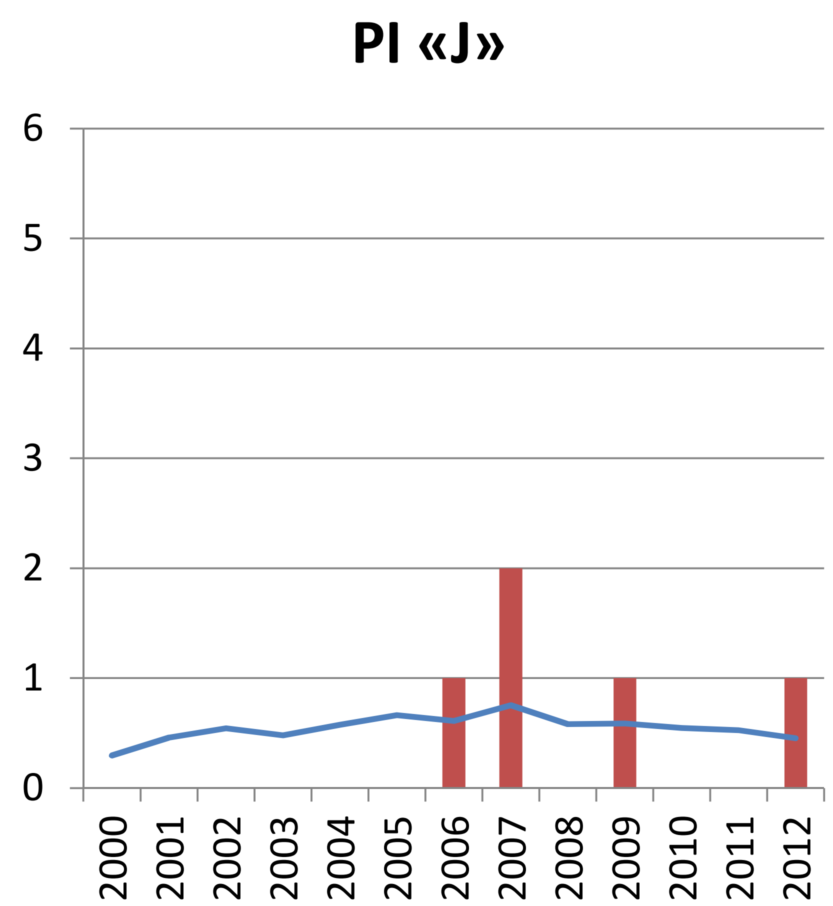
## PI «H» (EPO patents**)



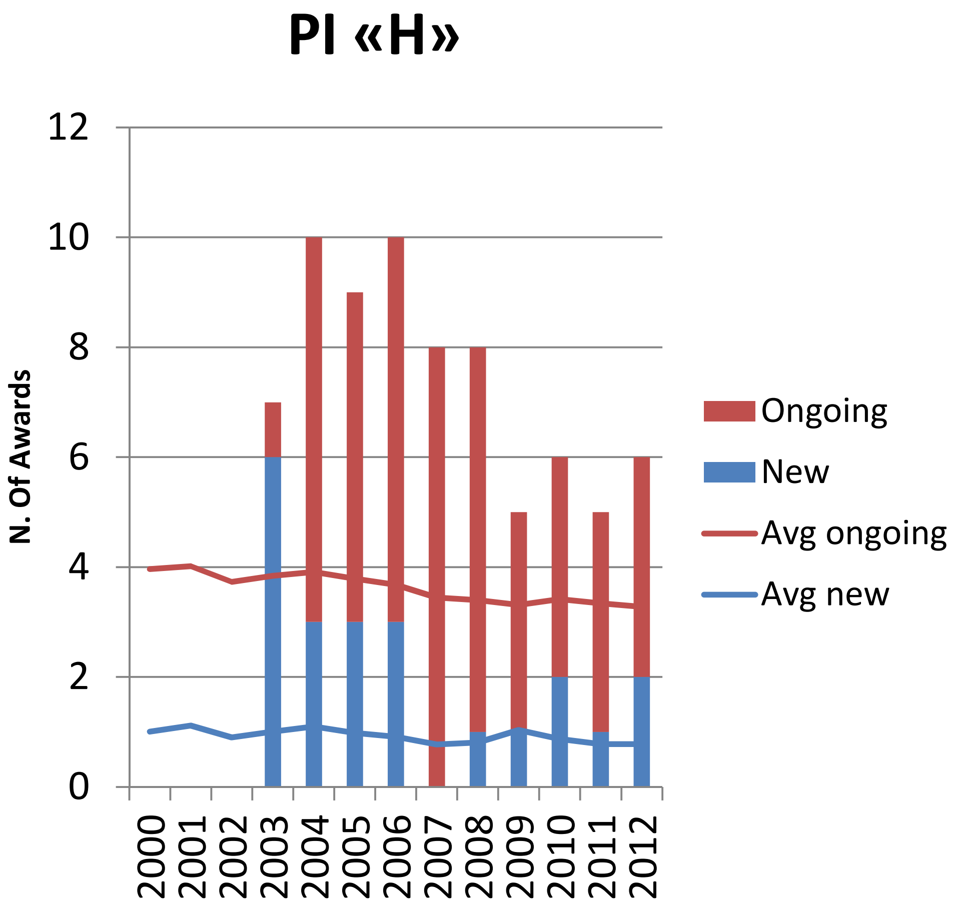

** patents by filing date

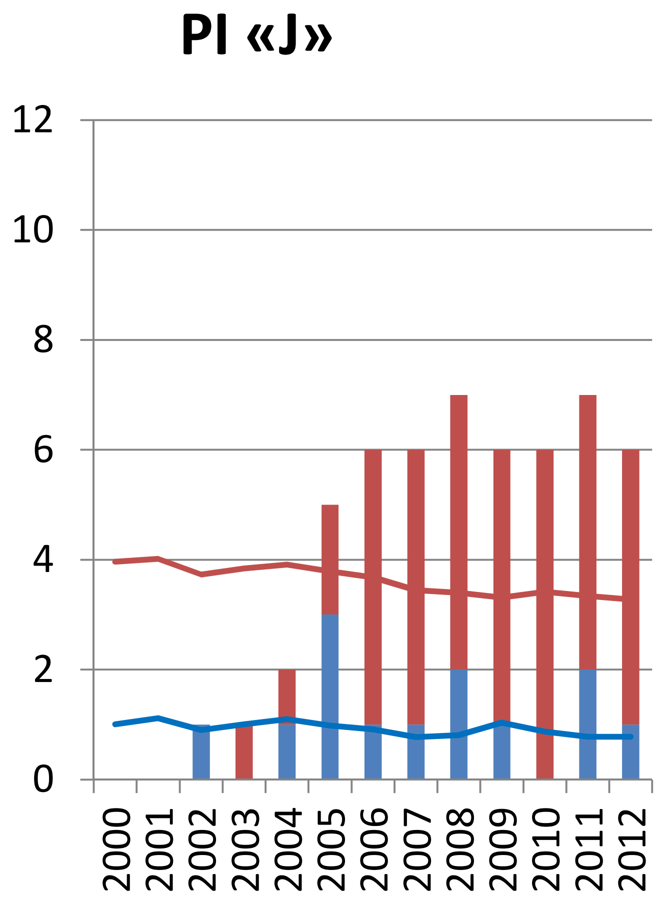
# Number of Ph.D. Theses Supervised



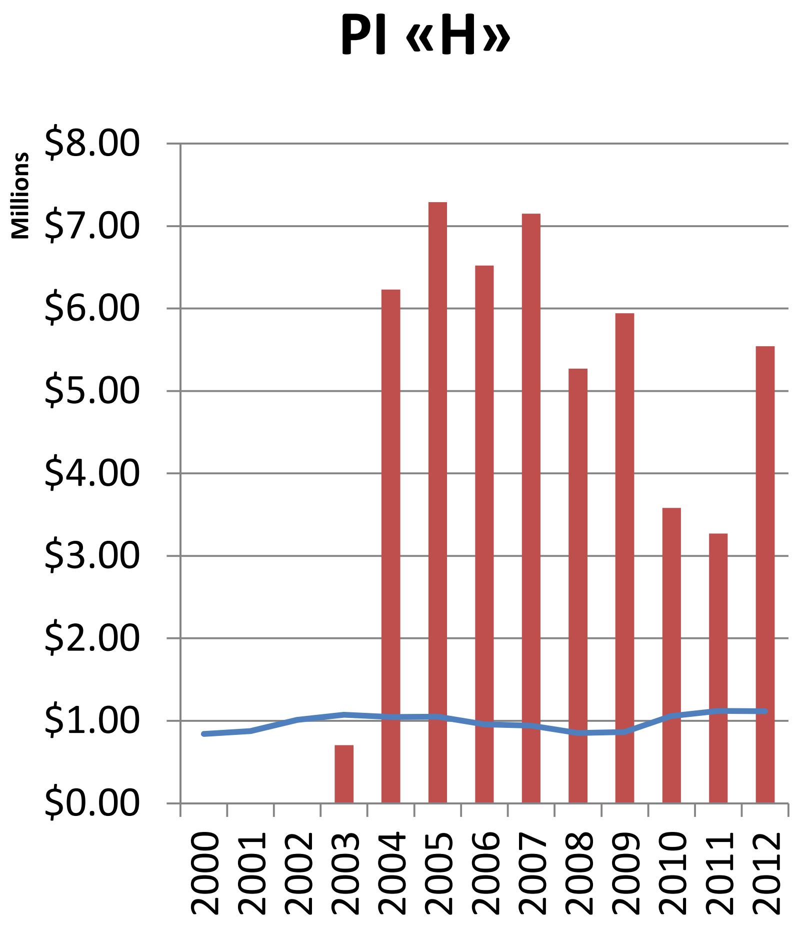



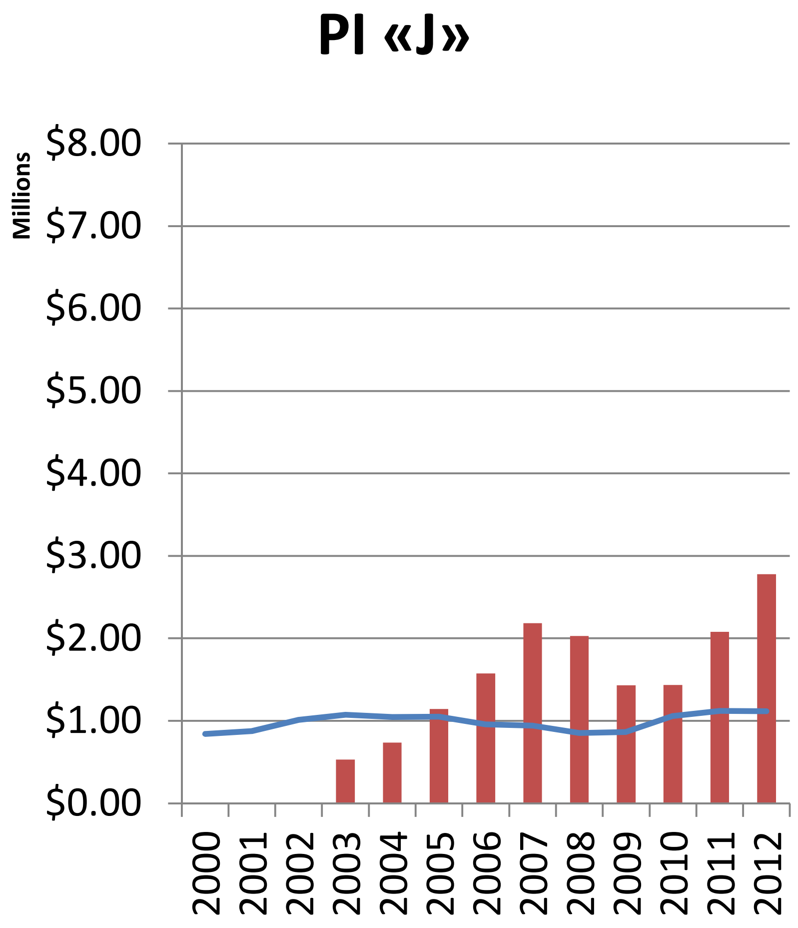
# Number of ongoing and new awards



# Flow of awards per year (millions)



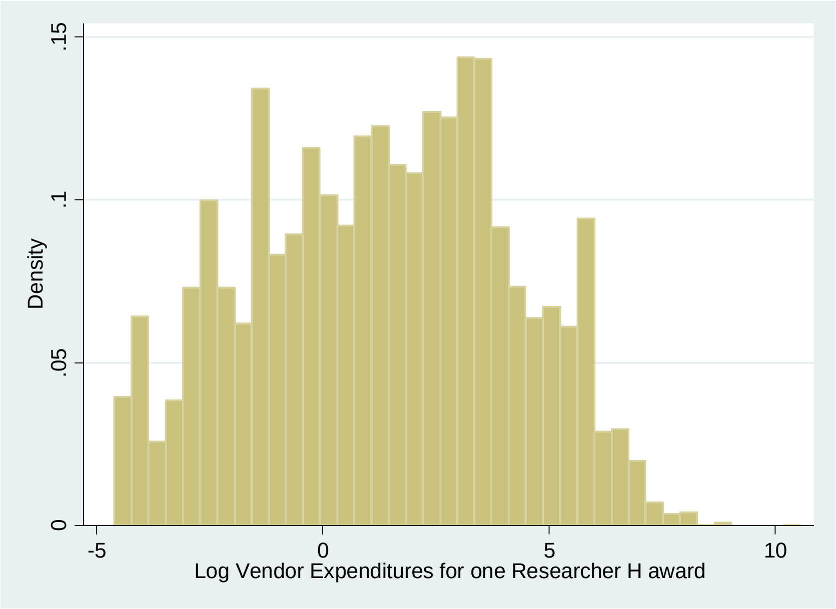



# Detailed vendor expenditure information for one award for Researcher J

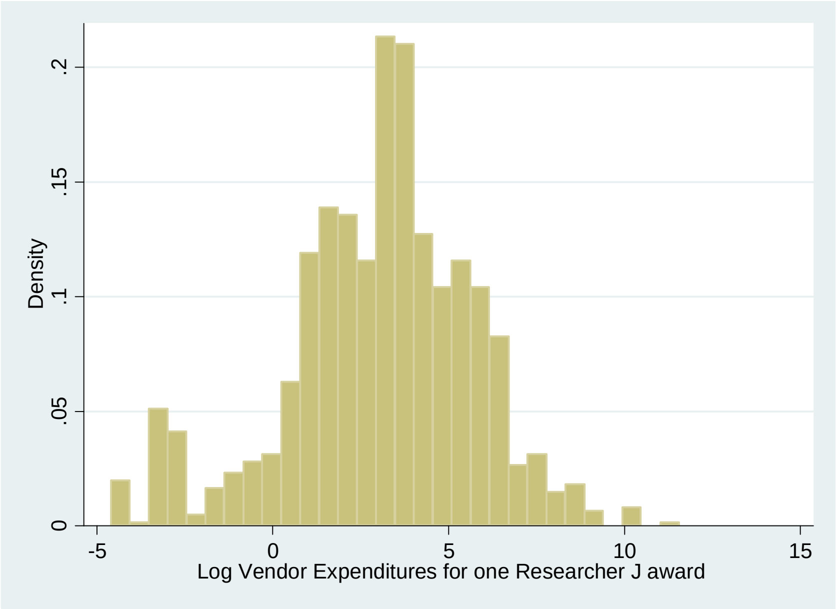
| Expenditure Distribution for One Award of Researcher J | | |
|---|---|---|
| Industry | Expenditures in dollars | Number of transactions |
| Other Professional Equipment and Supplies | 3386.36 | 121 |
| Rail transportation | 36 | 1 |
| Scenic and Sightseeing Transportation | 896.12 | 4 |
| Commercial Banking | 4616 | 2 |
| Testing Laboratories | 8312.92 | 100 |
| Pharmaceutical Preparation Manufacturing | 629.63 | 12 |
| Biological Product (except Diagnostic) Manufactures | 2480.45 | 37 |
| Electrometallurgical Ferroalloy Product | 189.8 | 8 |
| Electronic Computer Manufacturing | 6831.41 | 49 |
| Semiconductor and Related Device Manufactures | 3672.51 | 73 |
| Analytical Laboratory Instrument Manufactures | 61464.87 | 49 |
| Scheduled Passenger Air Transportation | 5892.79 | 19 |
| Passenger car rental | 1015.28 | 8 |
| Research and development in the physical science | 1654.88 | 38 |
| Colleges, Universities, and Professional | -110.88 | 1 |

# Distribution of log vendor expenditures

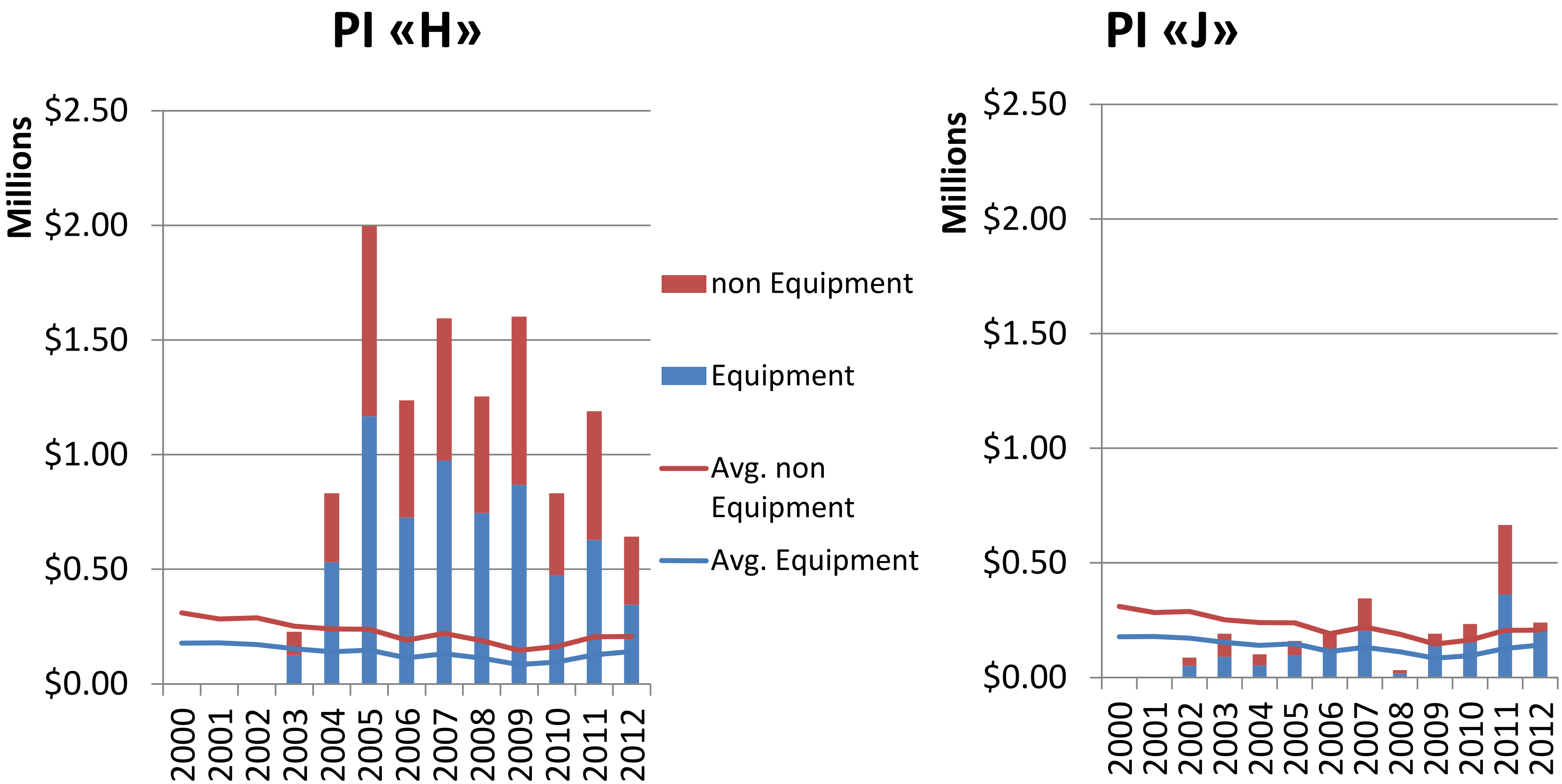
**PI «H»**




**PI «J»**



# Flow of Expenditures (millions)



# Descriptive Statistics (1)

## Personal Characteristics

| | N | mean | sd | Q1 | Q2 | Q3 |
|---|---|---|---|---|---|---|
| PhD Seniority | 2590 | 25.06 | 13.82 | 14 | 24 | 34 |

## Productivity measures

| | N | mean | sd | Q1 | Q2 | Q3 |
|---|---|---|---|---|---|---|
| log(1+pub) | 2590 | 1.13 | 0.92 | 0.00 | 1.10 | 1.95 |
| log(1+pat) | 2590 | 0.10 | 0.33 | 0.00 | 0.00 | 0.00 |
| log(1+phd) | 2590 | 0.33 | 0.47 | 0.00 | 0.00 | 0.69 |

## Labor, Equipment and Awards

| | N | mean | sd | Q1 | Q2 | Q3 |
|---|---|---|---|---|---|---|
| log(1+Labor) [2 years ma] | 2590 | 2.05 | 1.29 | 1.10 | 2.14 | 3.02 |
| share Post-Doc [2 years ma] | 2590 | 0.14 | 0.16 | 0.00 | 0.10 | 0.22 |
| share PhD [2 years ma] | 2590 | 0.28 | 0.24 | 0.00 | 0.25 | 0.44 |
| log(1+Equipment) [3 years ma] | 2590 | 8.60 | 4.43 | 7.50 | 10.06 | 11.71 |
| share Computer [3 years ma] | 2590 | 0.02 | 0.05 | 0.00 | 0.00 | 0.01 |
| share Optics [3 years ma] | 2590 | 0.01 | 0.05 | 0.00 | 0.00 | 0.01 |
| log(1+N. Awards) [3 years ma] | 2590 | 1.23 | 0.79 | 0.69 | 1.25 | 1.87 |
| log(1+F. Awards) [3 years ma] | 2590 | 10.49 | 5.40 | 10.42 | 12.79 | 13.96 |
| Share ARRA projects [3 years ma] | 2590 | 0.01 | 0.06 | 0 | 0 | 0 |

# Statistics (2)

**PhD university**

| | N | mean | sd | Q1 | Q2 | Q3 |
|---|---|---|---|---|---|---|
| Caltech | 2590 | 0.12 | 0.33 | 0 | 0 | 0 |
| MIT | 2590 | 0.11 | 0.31 | 0 | 0 | 0 |
| Harvard | 2590 | 0.06 | 0.24 | 0 | 0 | 0 |
| Princeton | 2590 | 0.07 | 0.26 | 0 | 0 | 0 |
| Stanford | 2590 | 0.09 | 0.29 | 0 | 0 | 0 |
| Berkeley | 2590 | 0.07 | 0.26 | 0 | 0 | 0 |
| Other institutions | 2590 | 0.48 | 0.50 | 0 | 0 | 1 |
| Total | | 1.00 | | | | |

**PhD field**

| | N | mean | sd | Q1 | Q2 | Q3 |
|---|---|---|---|---|---|---|
| Biology | 2590 | 0.06 | 0.24 | 0 | 0 | 0 |
| Physics | 2590 | 0.17 | 0.38 | 0 | 0 | 0 |
| Chemistry | 2590 | 0.07 | 0.26 | 0 | 0 | 0 |
| Engineering | 2590 | 0.15 | 0.36 | 0 | 0 | 0 |
| Mathematics | 2590 | 0.06 | 0.24 | 0 | 0 | 0 |
| Other topics | 2590 | 0.48 | 0.50 | 0 | 0 | 1 |
| Total | | 1.00 | | | | |

# Outline

- Current State: Economics of Science
  - What we know
  - How we know it
  - What we don’t know
  - The potential
- Caltech Project
  - Framework
  - Data Description
  - Empirical Analysis
- National and International Activities
- Next Steps

# Exploratory Regressions

$$\log(1+Y) =$$

$$\alpha \log(1+L) + \gamma_1 \left(\frac{L\ postdoc}{L}\right) + \gamma_2 \left(\frac{L\ PhD}{L}\right)$$

$$+\beta \log(1+K) + \delta_1 \left(\frac{K\ Computer}{K}\right) + \delta_2 \left(\frac{K\ Optics}{K}\right)$$

$$+ \sum_{t=2000}^{2012} \varphi_t\, T_t + \sum_{u=1}^{6} \varphi_u\, U_u + \sum_{f=1}^{5} \varphi_f\, F_f + \sigma\ Seniority$$

$$+ \ldots + Const + \varepsilon$$

# Labor productivity regression

| | (1) | (2) | (3) |
|---|---|---|---|
| | log(1+pub) | log(1+pat) | log(1+phd) |
| | | | |
| log(1+Labor) | 0.21*** | 0.056*** | 0.14*** |
| Share Post-Doc | 0.62*** | -0.027 | - |
| Share PhD | 0.29* | 0.034 | - |
| Calendar year dummies | yes | yes | yes |
| Constant | 0.060 | -0.018 | -0.58 |
| | | | |
| Observations | 2,590 | 2,590 | 2,590 |
| R-squared | 0.240 | 0.053 | 0.147 |
| Robust standard errors | | | |
| *** p<0.01, ** p<0.05, * p<0.1 | | | |

# Equipment productivity regression

| | (1) | (2) | (3) |
|---|---|---|---|
| | log(1+pub) | log(1+pat) | log(1+phd) |
| | | | |
| log(1+Equipment) | 0.066*** | 0.013*** | 0.033*** |
| Share Computer | 0.079 | 0.064 | -0.24 |
| Share Optics | 0.33 | 0.71** | 0.45** |
| Calendar year dummies | yes | yes | yes |
| Constant | 0.12 | -0.0058 | -0.030 |
| | | | |
| Observations | 2,590 | 2,590 | 2,590 |
| R-squared | 0.201 | 0.050 | 0.105 |
| Robust standard errors | | | |
| *** p<0.01, ** p<0.05, * p<0.1 | | | |

# Labor and Equipment productivity regression

| | (1) | (2) | (3) |
|---|---|---|---|
| | log(1+pub) | log(1+pat) | log(1+phd) |
| Labor | | | |
| Log(1+Labor) | 0.18*** | 0.051*** | 0.13*** |
| Share Post-Doc | 0.57** | -0.041 | - |
| Share PhD | 0.27* | 0.013 | - |
| Equipment | | | |
| Log(1+Equipment) | 0.015 | 0.0017 | 0.0041 |
| Share Computer | -0.11 | 0.00089 | -0.43 |
| Share Optics | 0.36 | 0.74** | 0.53** |
| Calendar year dummies | yes | yes | yes |
| Constant | 0.025 | -0.023 | -0.070* |
| | | | |
| Observations | 2,590 | 2,590 | 2,590 |
| R-squared | 0.243 | 0.065 | 0.153 |
| Robust standard errors | | | |
| *** p<0.01, ** p<0.05, * p<0.1 | | | |

# Award number and flow productivity regression

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | N. Awards | | | F. Awards | | |
| | log(1+pub) | log(1+pat) | log(1+phd) | log(1+pub) | log(1+pat) | log(1+phd) |
| | | | | | | |
| Log(1+N. Awards) / log(1+F. Awards) | 0.56*** | 0.10*** | 0.24*** | 0.072*** | 0.010*** | 0.029*** |
| Calendar year dummies | yes | yes | yes | yes | yes | yes |
| Constant | -0.039 | -0.024 | -0.062 | -0.083 | 0.0016 | -0.061 |
| | | | | | | |
| Observations | 2,590 | 2,590 | 2,590 | 2,590 | 2,590 | 2,590 |
| R-squared | 0.324 | 0.065 | 0.166 | 0.278 | 0.033 | 0.119 |
| Robust standard errors | | | | | | |
| *** p<0.01, ** p<0.05, * p<0.1 | | | | | | |

# Awards, Labor and Equipment productivity regression

| | (1) | (2) | (3) |
|---|---|---|---|
| VARIABLES | log(1+pub) | log(1+pat) | log(1+phd) |
| Awards | | | |
| Log(1+F. Awards) | 0.058*** | 0.0013 | 0.012*** |
| Labor | | | |
| log(1+Labor) | 0.096** | 0.049*** | 0.11*** |
| Share Post-Doc | 0.37 | -0.046 | - |
| Share PhD | 0.063 | 0.0085 | - |
| Equipment | | | |
| log(1+Equipment) | -0.011 | 0.0011 | -0.0019 |
| Share Computer | -0.15 | -0.000056 | -0.42 |
| Share Optics | 0.59 | 0.74** | 0.56** |
| Calendar year dummies | yes | yes | yes |
| Constant | -0.11 | -0.026 | -0.10*** |
| | | | |
| Observations | 2,590 | 2,590 | 2,590 |
| R-squared | 0.290 | 0.065 | 0.162 |
| Robust standard errors in parentheses | | | |
| *** p<0.01, ** p<0.05, * p<0.1 | | | |

# Awards, Labor, Equipment, and PI's Characteristics productivity regression ...

| | (1) | (2) | (3) |
|---|---|---|---|
| | log(1+pub) | log(1+pat) | log(1+phd) |
| Awards, Labor and Capital | | | |
| log(1+Award Amount) | 0.057*** | 0.0016 | 0.011*** |
| Share ARRA projects | 0.84** | -0.12 | -0.18 |
| log(1+Labor) | 0.11*** | 0.051*** | 0.11*** |
| Share post-doc | 0.26 | -0.073 | - |
| Share PhD | -0.12 | -0.027 | - |
| log(1+Equipment) | -0.015* | 0.00023 | -0.0022 |
| Share computer | -0.47 | -0.100 | -0.39 |
| Share optics | 0.055 | 0.64** | 0.38* |
| Personal Characteristics | | | |
| Seniority | -0.0082*** | -0.00044 | 0.00048 |
| | | | |
| Continue… | | | |
| Observations | 2,590 | 2,590 | 2,590 |
| R-squared | 0.361 | 0.097 | 0.200 |

# … Awards, Labor, Equipment, and PI's Characteristics productivity regression

| | (1) | (2) | (3) |
|---|---|---|---|
| | log(1+pub) | log(1+pat) | log(1+phd) |
| PhD field | | | |
| Biology | 0.070 | 0.036 | -0.049 |
| Physics | 0.24*** | -0.033 | 0.012 |
| Chemistry | 0.41*** | 0.074 | 0.19** |
| Engineering | 0.48*** | 0.098 | 0.23*** |
| Mathematics | -0.011 | 0.010 | 0.043 |
| PhD institution | | | |
| Caltech | 0.28*** | 0.019 | 0.045 |
| MIT | 0.083 | 0.071 | -0.012 |
| Harvard | -0.014 | -0.042 | -0.028 |
| Princeton | 0.11 | -0.057* | 0.067 |
| Stanford | 0.11 | -0.048 | -0.0020 |
| Berkeley | 0.11 | 0.032 | -0.078 |
| | | | |
| Constant | -0.038 | -0.017 | -0.14*** |
| Observations | 2,590 | 2,590 | 2,590 |
| R-squared | 0.361 | 0.097 | 0.200 |

# Outline

- Current State: Economics of Science
  - What we know
  - How we know it
  - What we don't know
  - The potential
- Caltech Project
  - Framework
  - Data Description
  - Empirical Analysis
- National and International Activities
- Next Steps

# CIC Activity

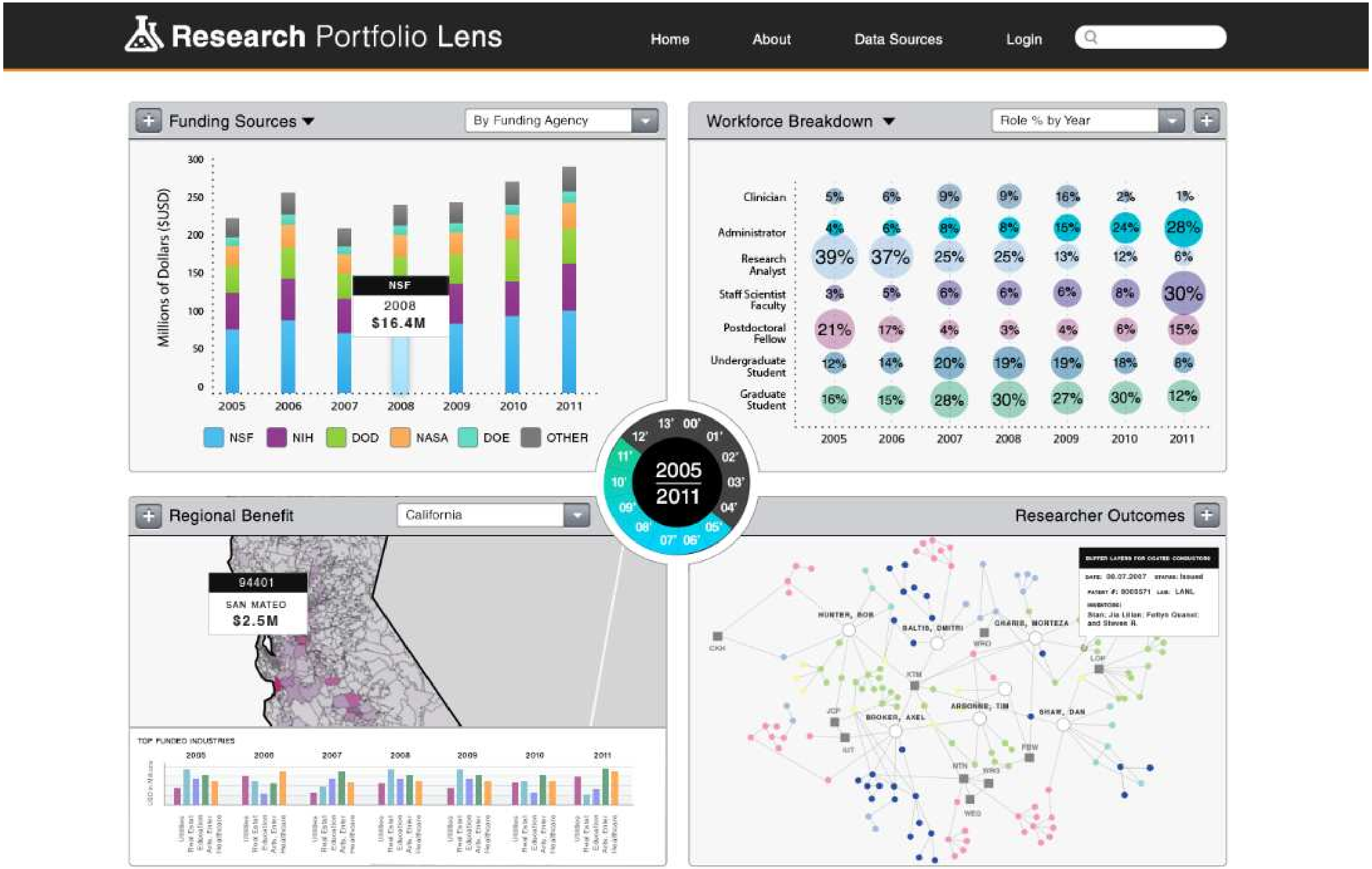
- Roy Weiss, University of Chicago, lead instigator
- March workshop hosted by CIC
  - CIC VPs for Research
  - CIC Science of Science Policy Researchers
  - NIH, NSF, USDA, Sloan Foundation, AAMC, AAU and APLU as observers
- All institutions agreed to collaborate in data (in secure site) and analysis
  - Projects now started or proposed in (1) economic impact; (2) food safety and security (3) postdoc outcomes

# National (CIC) Activity

## Range of Studies

| | Zoom Out | | | | Zoom In |
|---|---|---|---|---|---|
| Analysis of: | Entire scientific enterprise | Scientific fields | Entire research institutions and funders | Specific fields at specific institutions | Specific labs, researchers |
| Useful for: | Government and institutions to justify and set level of science investments | Government to justify and allocate investments to fields | Institutions and funders to document performance | Identify best-practices, target investments | Micro-benchmark performance, identify underexploited opportunities |
| Method: | Econometric analysis | Econometric analysis | Bibliometric, Text, + Econometric | Bibliometric, Text, + Econometric | Bibliometric, Text, + Econometric |
| Status: | In progress | In progress | Planned | Planned | Planned |

# Engaging research institutions



# International Context

STAR METRICS (USA/White House, NSF, NIH, DOE, USDA, EPA)

ASTRA (Australia)

HELIOS (France)

CAELIS (Czech Republic)

NORDSJTERNEN (Norway)

STELLAR (Germany)

TRICS (UK)

TBD (SPAIN)

# Outline

- Current State: Economics of Science
  - What we know
  - How we know it
  - What we don’t know
  - The potential
- Caltech Project
  - Framework
  - Data Description
  - Empirical Analysis
- National and International Activities
- Next Steps

# Next Steps: Caltech project

- Data
  - Extensive data cleaning and variable construction
  - Measures of space, placement
  - Build open source/big data platform for multiple organizations
- Presentations
  - Academic: AAAS, Southern, Paris September
  - Community: APLU/Council of Graduate Deans/AAU/CIC
- Publication
  - Science/Nature; PNAS; ……..

# Next Steps: Data Infrastructure

- Follow on from Sloan workshop in Boston July 14; White paper end of August
- Work with Big Data researchers (Jeff Hammerbacher in particular in August)
- Engage funders, big data and science policy communities in fall