The Economics of Science

an empiricist's view

Kyle R. Myers

Innovation Research Boot Camp, Summer 2024

Scientists as choosers

demand: preferences and adjustment costs

- **Stern**. "Do scientists pay to be scientists?"

  Management Science 50, no. 6 (2004): 835-853.

- **Myers**. "The elasticity of science."

  American Economic Journal: Applied Economics 12, no. 4 (2020): 103-134.

- **Acemoglu**. "Diversity and technological progress."

  The Rate and Direction of Inventive Activity Revisited (2011). U. Chicago Press, 319-356.

Scientists as producers

supply: the basic—applied spectrum

- **Azoulay, Li, Graff Zivin, & Sampat**. "Public R&D Investment and Private Sector Patenting: Evidence from NIH Funding Rules."

  The Review of Economic Studies 86, no. 1 (2019): 117-152.

- **Myers & Lanahan**. "Estimating Spillovers from Publicly-Funded R&D: Evidence from the US Department of Energy."

  American Economic Review 112, no. 7 (2022): 2393-2423.

- **Bloom, Schankerman, & Van Reenen**. "Identifying Technology Spillovers and Product Market Rivalry."

  Econometrica 81, no. 4 (2013): 1347-1393.

Aside:
Estimating (Innovation) Production Functions

Exponential Production Functions

a simple starting point

Structural prod. func.: $\log\left(Y_{it}\right) = a + \beta \log(X_{it}) + \omega_{it} + \epsilon_{it}$

Objective func.: output maximization subject to budget

Optimal investment policy: $X_{it}^{*} = i(\omega, \ldots)$

A good research design requires understanding $i(\ldots)$!

Stocks and Flows

$$\log\left(Y_{it}\right) = a + \beta_1 \log(X_{it}) + \beta_2 \log(X_{i(t-1)}) + \omega_{it} + \epsilon_{it}$$

- **Zvi Griliches**: “knowledge stock” = $(1-\delta)^0 X_{it} + (1-\delta)^1 X_{i(t-1)} + (1-\delta)^2 X_{i(t-2)} + \ldots$
  - Depreciates at a rate $\delta < 1$: some R&D outputs are persistent knowledge
  - Depreciates at a rate $\delta \geq 0$: some R&D inputs are variable costs
  - *"Issues in assessing the contribution of research and development to productivity growth." The Bell Journal of Economics (1979)*
- **Bronwyn Hall**: nitty-gritty (but important!) empirics of R&D stocks
  - *“Measuring the Returns to R&D: The Depreciation Problem.” NBER Working Paper (2007)*

Production Functions and Fixed Effects

$$\log\left(Y_{it}\right) = a_i + \beta_1 \log(X_{it}) + \omega_{it} + \epsilon_{it}$$

- **Griliches & Mairesse**: "Production Functions: The Search for Identification." (1995)

"***Researchers, in trying to evade the simultaneity problem…***"

[by using panel data, including producer-fixed effects, and assuming that X_{it} and ω_{it} are independent conditional on a_i]

"***…have shifted to the use of thinner and thinner slices of data…***"

[identifying β only via variation from $\log(X_{it}) - \alpha_i$]

exacerbating other problems and misspecifications."

Empirical Industrial Organization:
Models, Methods, and Applications

Victor Aguirregabiria

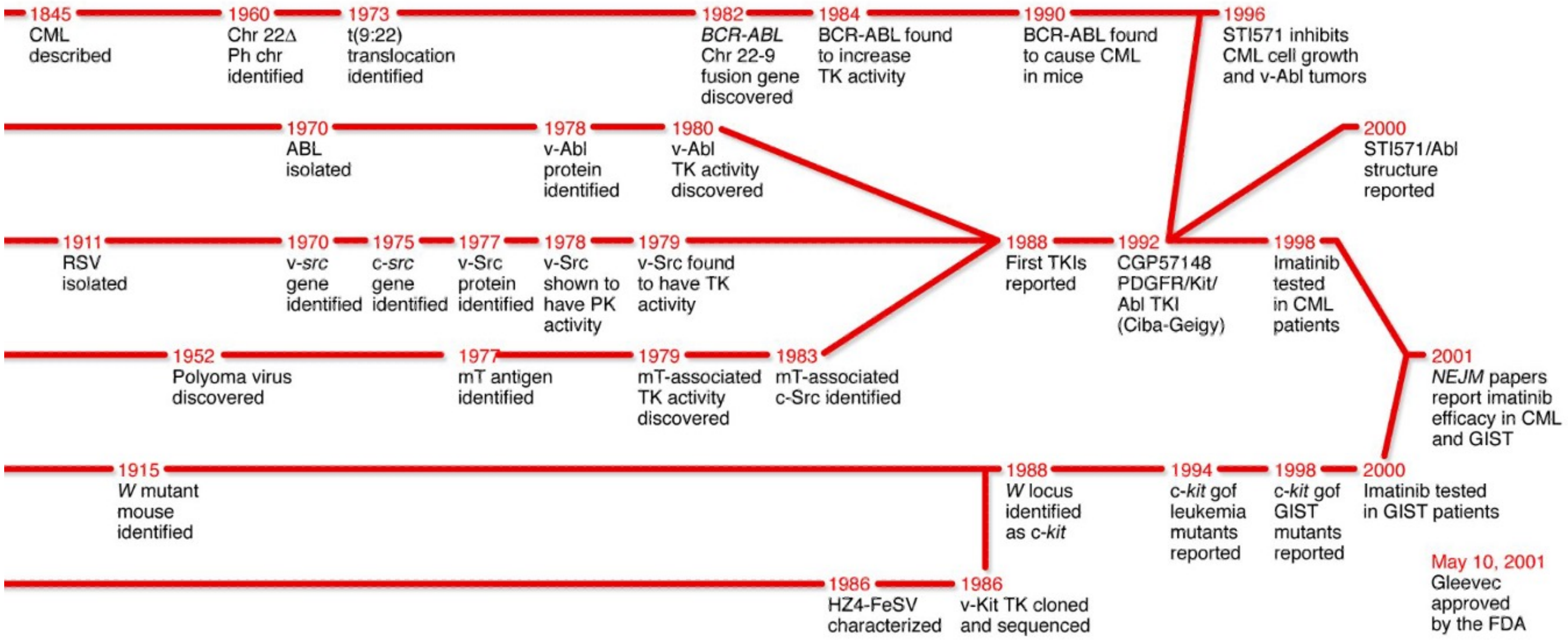
Public R&D Investment and Private Sector Patenting: Evidence from NIH Funding Rules

Azoulay, Li, Graff Zivin, & Sampat

The Review of Economic Studies 86, no. 1 (2019): 117-152

Is Science (eventually) Valuable?

the long road of Gleevec



Unit of analysis: D(isease)**-S**(cience)**-T**(ime)

- No scientist does research “on cancer”
  - Work involves a science area and a disease application (e.g., cell signaling in cancer)
- Here, research area = disease-science area for a given year
  - Work that uses similar **tools / biological-pathways** (science) to make progress towards treatments for the same **illness, injury, or disorder** (disease) in the same **year** (time)
- Advantages
  - Allows a policy-relevant question: what happens if we provide more funding for a disease-science area? (e.g. genetic basis of Alzheimer's)
  - D-S-T are not explicit units of funding for NIH administration (which will help with identification)

Defining each D-S-T

Funding	Evaluation
NIAAA National Institute on Alcohol Abuse and Alcoholism National Heart Lung and Blood Institute NIBIB NATIONAL CANCER INSTITUTE NIDA NATIONAL INSTITUTE ON DRUG ABUSE NIMH National Institute of Mental Health NLM	NIH Center for Scientific Review 

Defining each D-S-T

- Defining “**diseases**”:
  - NIH consists of 27 disease(ish)-focused Institutes/Centers
  - A grant application must report its disease area to be funded
- Defining “**science**”:
  - Grant review happens in 180 science(ish)-focused “study sections”
  - A grant application must specify its science area to be evaluated
- Defining “**time**”:
  - Fiscal years

Empirics

$$Patents_{???} = a + \beta Funding_{dst} + \epsilon_{dst}$$

- *Where to look for outcomes? (because patents aren't explicitly assigned to DSTs)*
  - It is hard to know a priori what scientific results are relevant for a patent
  - Link grants to patents via:
    - Paper trail: acknowledgements — NIH funding directly used
    - Paper trail: citations — patent cites a paper that NIH funded
    - ``Nearby'' in disease-science space (i.e., using similar language)

Finding Patents
connected to NIH investments

- **Direct acknowledgment**: # patents by NIH-funded researchers
  - Grant → Patent
  - *Answers*: Does the NIH directly fund patentable research?
- **Citation-linked**: # patents citing NIH-funded research
  - Grant → Publication → Patent
  - *Answers*: Does the NIH fund research that is directly useful to inventors?
- **``Near-by''**: # patents intellectually related to an NIH funding area
  - Grant → Publication → Related Publication → Patent
  - *Answers*: Does the NIH fund research that is indirectly useful to inventors?

Identification

$$Patents_{d(\delta)s(\sigma)t(\tau)} = a + \beta Funding_{dst} + \epsilon_{dst}$$

- **Concern**: $Funding_{dst}$ may be correlated with ϵ_{dst}
- **Approach 1**: Fixed effects
  - Assumption: $\epsilon_{dst} = (FE_d \times FE_s) + (FE_d \times FE_t) + (FE_s \times FE_t) + \mu_{dst}$
  - Scientists and the NIH (may) know everything, except for μ_{dst}
- **Approach 2**: Instrumental variable — “windfall” funding due to funding rules
  - DST funding is made up of funding for individual grants.
  - Grant applications are given cardinal scores, but funded on the basis of ordinal scores.
  - Instrument $Funding_{dst}$ with funding for the subset of grants funded for this reason

“Windfall” Funding

Cell Signaling Study Section			Tumor Physiology Study Section		
Rank	Disease	Raw Score	Rank	Disease	Raw Score
1	Cancer	10	1	Cancer	8.2
2	Diabetes	9.8	2	Cancer	8.1
3	Cancer	9.2	3	Cancer	7.6
4	Cancer	9.1	4	Cancer	6.4
5	Cancer	8.2	5	Cancer	5.4
6	Diabetes	7.6	6	Diabetes	5.2
7	Cancer	7.6	7	Diabetes	4.8
8	Diabetes	7.5	8	Diabetes	4.4

Main Results: NIH $ →

- 30% of NIH grants produce research that is cited by a private sector patent
- $10 million of NIH funding → 2.3 more industry patents
- NIH funding increases overall firm R&D investment
    - Increased firm patenting in one area is not offset by declines in another; rather, both appear to increase
- $1 dollar in NIH funding → $0.4 to $1.7 in PDV of drug revenue
- Disease spillovers are large
    - Half of all patents generated by additional NIH investments are for diseases different from the one intended

Estimating Spillovers from Publicly-Funded R&D: Evidence from the US Department of Energy

Myers & Lanahan

American Economic Review 112, no. 7 (2022): 2393-2423

Motivation: R&D spillovers
(ex-post rationalization of being a 1st-year AP)

- In theory, positive externalities from science $\Rightarrow$ gov't invests in science
- But little (micro) evidence on how big and "where" those externalities might be
  - *Azoulay, Li, Graff Zivin, & Sampat. "Public R&D Investment and Private Sector Patenting." The Review of Economic Studies (2019).* [basic, biomed.]
  - *Bloom, Schankerman, & Van Reenen. "Identifying Technology Spillovers and Product Market Rivalry." (2013).* [corporate R&D tax credits]

What actually happened …

Small business R&D + Energy sector

key recent work

Financing Innovation: Evidence from R&D Grants

Sabrina T. Howell

AMERICAN ECONOMIC REVIEW
VOL. 107, NO. 4, APRIL 2017
(pp. 1136–64)

Download Full Text PDF

Article Information

Abstract

Governments regularly subsidize new ventures to spur innovation. This paper conducts the first large-sample, quasi-experimental evaluation of R&D subsidies. I use data on ranked applicants to the US Department of Energy's SBIR grant program. An early-stage award approximately doubles the probability that a firm receives subsequent venture capital and has large, positive impacts on patenting and revenue. These effects are stronger for more financially constrained firms. Certification, where the award contains information about firm quality, likely does not explain the grant effect. Instead, the grants are useful because they fund technology prototyping.

October 01 2018

Approximating Exogenous Variation in R&D: Evidence from the Kentucky and North Carolina SBIR State Match Programs

Lauren Lanahan, Maryann P. Feldman

Author and Article Information

The Review of Economics and Statistics (2018) 100 (4): 740–752.
https://doi.org/10.1162/rest_a_00681 Article history

Cite Permissions Share Views

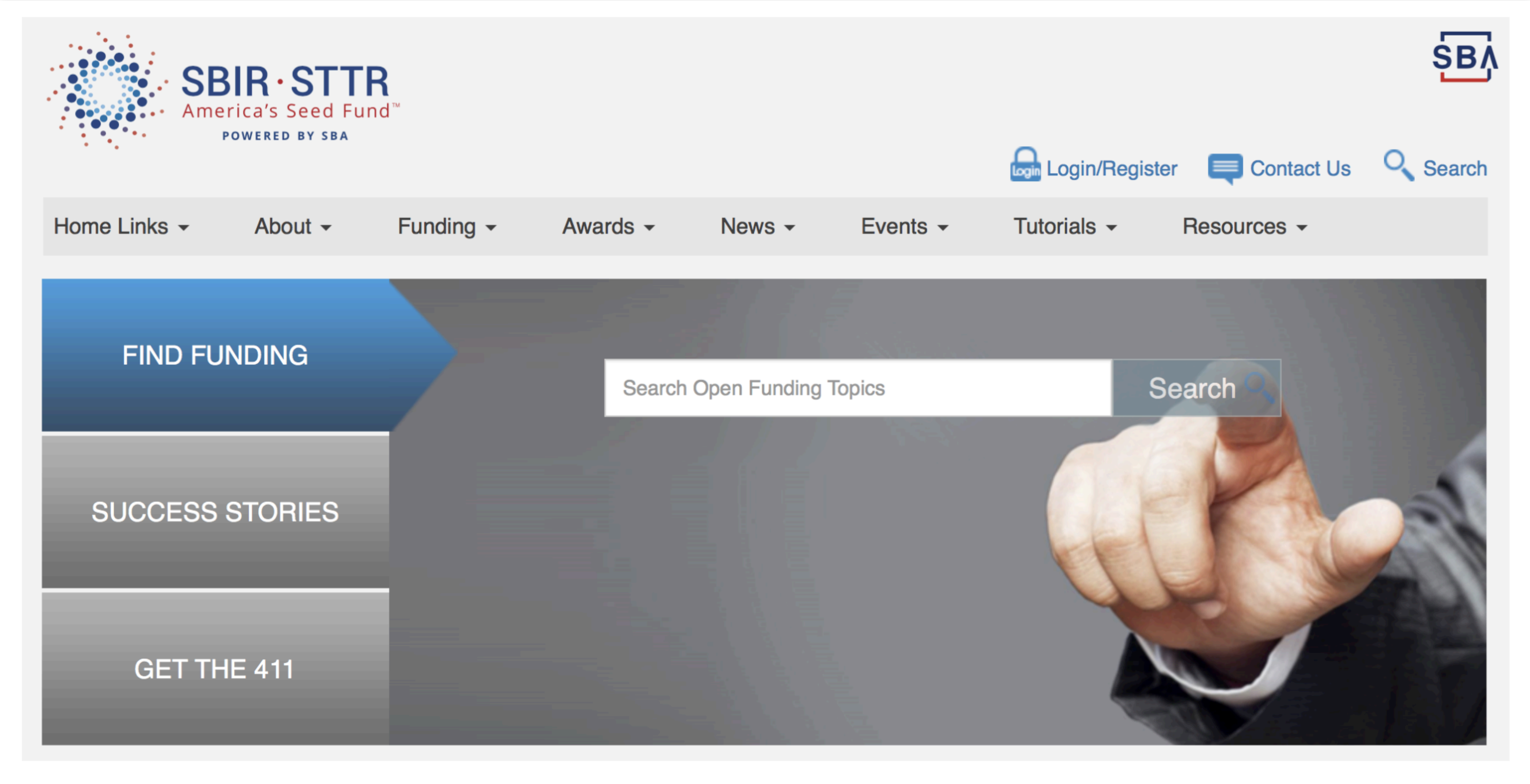
Abstract

This paper exploits policy discontinuities at U.S. state borders to examine the effect of R&D investments on innovative projects. We examine the Small Business Innovation Research (SBIR) State Match program, which offers noncompetitive grants to federally awarded SBIR Phase I projects that are eligible to compete for Phase II. Results from SBIR activity (2002–2010) indicate heterogeneous treatment effects. Notably, the positive differential effects are moderated by firms within the science and health fields and with less previous SBIR success. The State Match effectively stabilized Phase II trends in contrast to neighboring states that experienced greater declines from the concurrent recession.

SBIR at the DOE

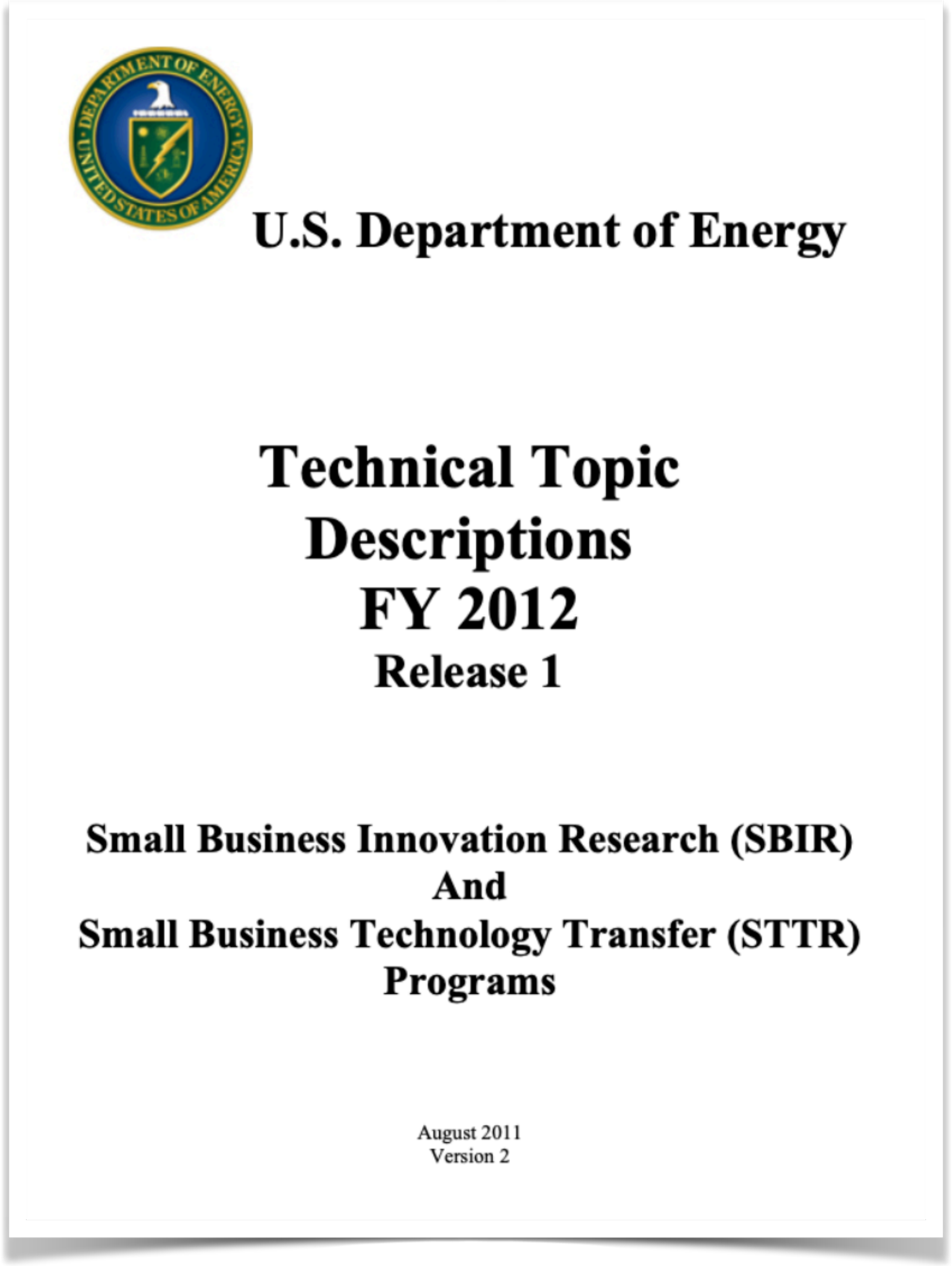
(and lots of other public science programs)

- **Small business**: for-profit company with $\leq$ 500 employees
- **SBIR award**: ~$150K (Phase I) & ~$1.5M (Phase II) grants for R&D



SBIR at the DOE: Targeted investment

(and lots of other public science programs)


U.S. Department of Energy

Technical Topic
Descriptions
FY 2012
Release 1

Small Business Innovation Research (SBIR)
And
Small Business Technology Transfer (STTR)
Programs

August 2011
Version 2

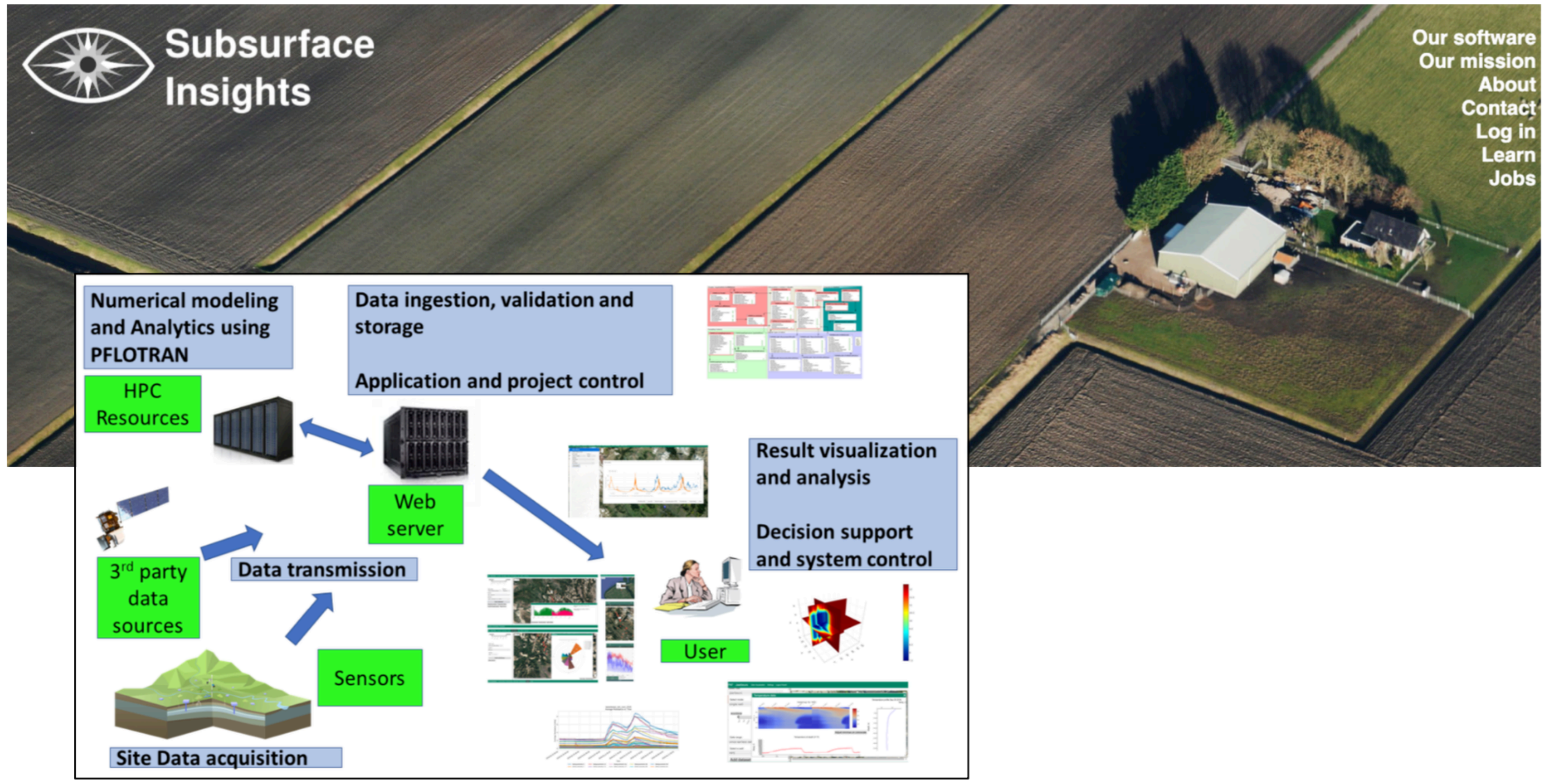
20. TECHNOLOGIES FOR SUBSURFACE CHARACTERIZATION AND MONITORING (PHASE I, $150,000/PHASE II, $1,000,000)

In support of the Department of Energy's (DOE's) secure and sustainable energy mission the Office of Biological and Environmental Research seeks to advance fundamental understanding of coupled biogeochemical processes in complex subsurface environments to enable systems-level prediction and decision support. This basic scientific understanding is applicable to a wide range of DOE relevant energy and environmental challenges including:

- Cleanup of contaminants and stewardship of former weapons production sites
- Underground storage of spent nuclear fuel
- Carbon cycling and sequestration in the environment
- Nutrient cycling in the environment in support of sustainable biofuel development
- Fossil fuel processing and recovery from the deep subsurface.

Subsurface Insights, Inc.
A case study

- **Initial location**: New Hampshire
  - **Later**: projects across U.S. & Europe
- **Initial SBIR topic**: monitor contamination zones
  - **Later**: supplying aquifer thermal energy storage companies



SBIR at the DOE: Targeted investment

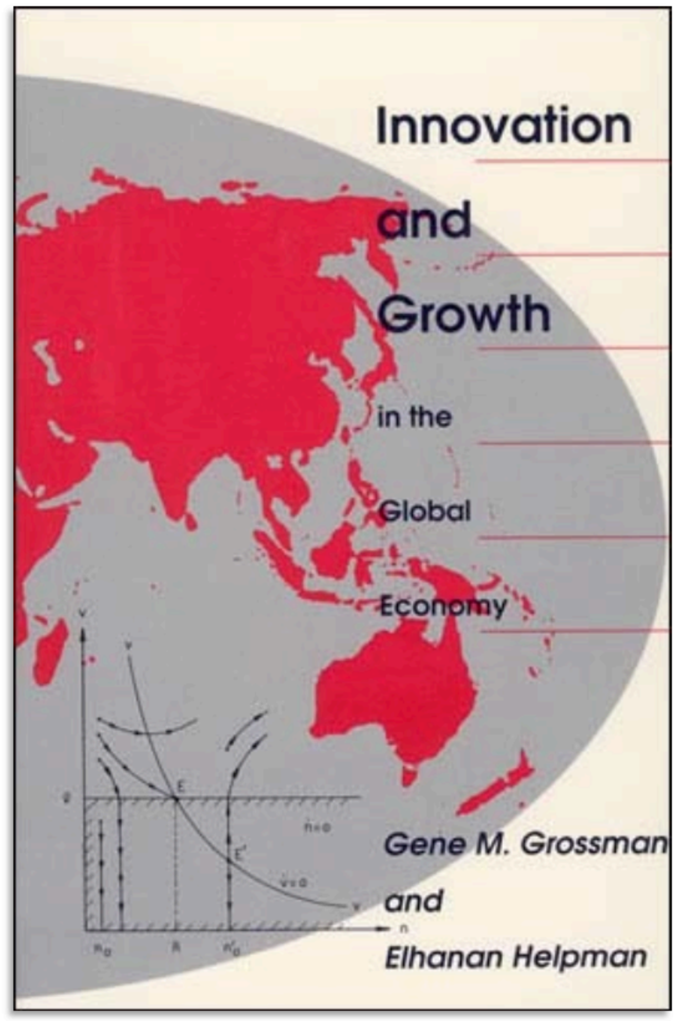
(and lots of other public science programs)


THE HILL

Energy isn't a horse race, and the government doesn't pick winners

BY DREW BOND, OPINION CONTRIBUTOR — 02/06/18 01:30 PM EST
THE VIEWS EXPRESSED BY CONTRIBUTORS ARE THEIR OWN AND NOT THE VIEW OF THE HILL

36 COMMENTS




ITIF | INFORMATION TECHNOLOGY & INNOVATION FOUNDATION

ISSUES | REGIONS | PUBLICATIONS | EVENTS | NEWS ROOM | ABOUT | DONATE

Energy Innovation in the FY 2021 Budget: Congress Should Lead

By Colin Cunliff | March 30, 2020

Lawmakers should make energy innovation a national imperative, and at least double U.S. investment in clean energy RD&D by 2025.


CNBC MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV

POLITICS

The government as venture capitalist

PUBLISHED TUE, JUN 24 2014-6:10 PM EDT

Laura Tyson, former chair of the US President's Council of Economic Advisers

SHARE

Empirical model

patent production function

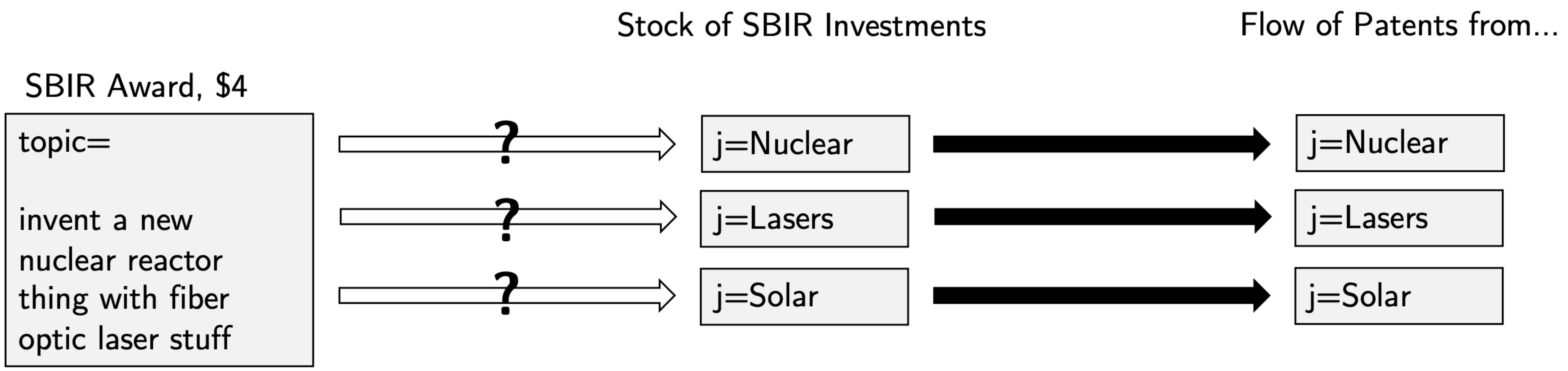
For each area of technology-space j in year t :

$$\mathbf{E}[\, y_{jt} \,|\, K_{jt} \,;\, \tau_t \,,\, \omega_{jt} \,] = \exp\left(\, \log(K_{jt}) + \tau_t + \omega_{jt} \, \right)$$

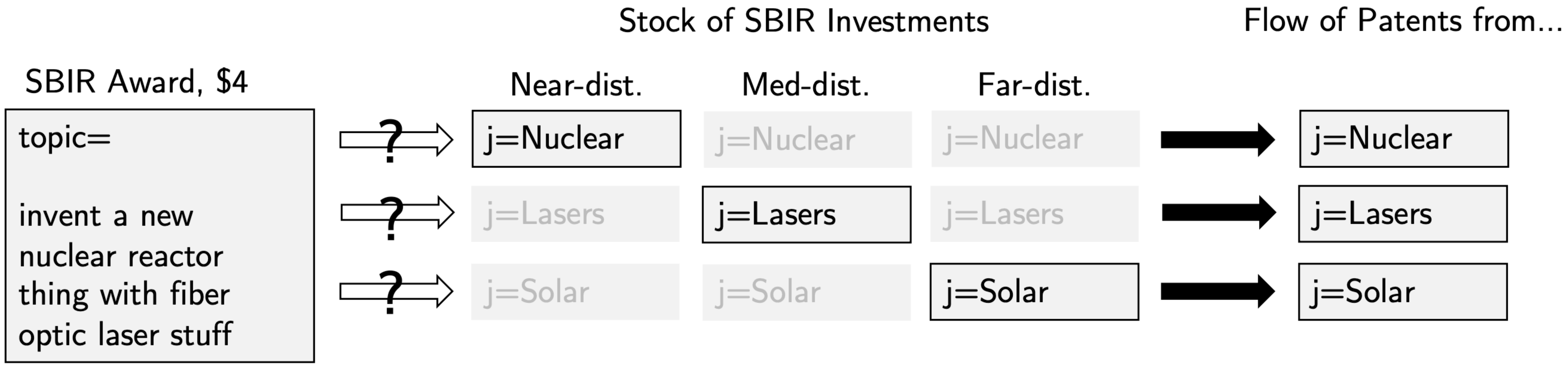
- y_{jt} — flow of patents in that space-year
- K_{jt} — stock of prior DOE SBIR $ in that space up until and including that year
- τ_t — aggregate trends
- ω_{jt} — unobservable supply and/or demand shocks in that space year

Empirical model

patent production function



Empirical model
patent production function



Mapping investments (SBIR $) to technology-space (CPC codes)

text-similarity

Step 1: "Read" FOA Topics
i.e. 2005 DOE Release 1, Topic #1

1. ADVANCED POWER ELECTRONICS FOR ENERGY STORAGE, TRANSMISSION, AND DISTRIBUTION APPLICATIONS

Power electronic conversion systems (PCS) constitute major cost elements and reliability issues in most distributed generation and energy storage systems. As these systems move to higher power levels, it is desirable to improve the functionality and manufacturability of the power conversion systems. Several paths to improvement are possible. Moving from silicon to silicon-carbide based devices has the potential to increase power rating and switching frequency while replacing electrolytic capacitors with other components offers the potential of significantly increasing the reliability of these devices.

Step 2: Find "similar" Patents

Patent #6891355
Issued: May 10, 2005 Filed: Nov. 14, 2002

Patent #6885170
Issued: April 26, 2005 Filed: Oct. 2, 2002
Title: Connection-switched capacitor storage system

Abstract: A connection-switched capacitor storage system comprises plural capacitors, parallel monitors connected with the capacitors, respectively, switches for switching the connections ...

Inventors: Okamura; Michio et al.
Assignee: Advanced Capacitor Technologies, Inc. et al.

CPC Classes: H02J7

Step 3: "Map" Classes to FOA Topics per Patent Similarity

1. H01L2924: methods for connecting or disconnecting semiconductor or solid-state bodies
2. H01L28: passive two-terminal components without a potential-jump or surface barrier for integrated circuits
3. H01L27: devices consisting of a plurality of semiconductor or other solid-state components formed in or on a common substrate

Mapping investments (SBIR $) to technology-space (CPC codes)

face validity of text-similarity mapping

Figure A.2: FOA Example #1–Solar Energy

(a) FOA Text

2. ADVANCED SOLAR TECHNOLOGIES

Solar energy is our largest energy resource and can provide clean, sustainable energy supplies, including electricity, fuels, and thermal energy. The President's economic recovery package emphasized solar energy, among others, as a key element in combating global climate change. However, the cost-effective capture of the enormous solar resource is problematic. This topic seeks to develop novel, commercially feasible, solar systems and production techniques.

Grant applications submitted in response to this topic should: (1) include a review of the state-of-the-art of the technology and application being targeted; (2) provide a detailed evaluation of the proposed technology and place it in the context of the current state-of-the-art in terms of lifecycle cost, reliability, and other key performance measures; (3) analyze the proposed technology development process, the pathway to commercialization, the large potential markets it will serve, and the attendant potential public benefits that would accrue; and (4) address the ease of implementation of the new technology.

Phase I should include (1) a preliminary design; (2) a characterization of laboratory-scale devices using the best measurements available, including a description of the measurement methods; and (3) a road map with major milestones, leading to a production model of a system that would be built in Phase II. In Phase II, devices suitable for near-commercial applications must be built and tested, and issues associated with manufacturing the units in large volumes at a competitive price must be addressed.

Grant applications are sought in the following subtopics:

a. Manufacturing Tools for Reliability Testing—Grant applications are sought for the development of tools that can be used to conduct reliability testing in PV module manufacturing environments. For example, tools such as light soaking equipment are used to prepare modules or components for accelerated lifetime testing, which is frequently conducted in-house at the module manufacturing facility or by service companies before sending for official third party certification. New tools are needed for the testing of components (e.g., modules, inverters) or subcomponents (e.g., cells, microinverters, individual layers of a module), and should combine high performance, low cost, and a small floor footprint.

Questions – contact: Alec Bulawka (Alec.Bulawka@ee.doe.gov)
James Kern (James.Kern@ee.doe.gov)

b. Module and System Manufacturing Metrology and Process Control—The rapid scale-up of the manufacturing of photovoltaics, particularly for new thin-film technologies, is challenging the possibility of using conventional technologies to make real-time non-destructive measurements of material characteristics in high-volume, high-production-rate environments and then using this information to implement real-time process control of the manufacturing process. Therefore, grant applications are sought for the development of novel, advanced, real-time non-destructive materials characterization tools for use in high-volume manufacturing lines for photovoltaic systems.

Questions – contact: Alec Bulawka (Alec.Bulawka@ee.doe.gov)
James Kern (James.Kern@ee.doe.gov)

c. Photovoltaics (PV) System Diagnostic Tools—The current rapid growth of the PV industry has led to diverse and innovative product designs, which frequently require non-traditional tests for reliability and performance. Examples of these non-traditional tests include performance testing and tracking requirements for concentrating PV modules, and software-based system diagnostic tools. Grant applications are sought for innovative methods to monitor PV system and component performance, in order to identify failures and loss mechanisms and to minimize system down time. Approaches of interest include the development of diagnostic tools that are process-oriented and internal to the system components, or those that can be integrated – i.e, "piggy-backed" – through ancillary application.

Questions – contact: Alec Bulawka (Alec.Bulawka@ee.doe.gov)
James Kern (James.Kern@ee.doe.gov)

(b) Titles of Relevant CPC Classes

Technology Distance ptile.	Example CPC Titles
1	Apparatus for processing exposed photographic materials Generation of electric power by conversion of infra-red radiation, visible light or ultraviolet light Plasma technique; production of accelerated electrically-charged particles
10	Electric heating; electric lighting Static electricity; naturally-occurring electricity Cyclically operating valves for machines or engines
20	Cranes; load-engaging elements or devices for cranes Locomotives; motor railcars Wireless communication networks

Notes: Topic #2 from the FY2010 Release 1 Funding Opportunity Announcement.

(a) FOA Text

4. GEOTHERMAL ENERGY TECHNOLOGY DEVELOPMENT

This topic is focused on the development and innovation required to achieve technical and commercial feasibility of EGS. Because of the complexity of these systems, grant applications are expected to focus on a component or supporting technology of EGS development that would enable improvements to the overall system. The unique function and innovation of the targeted subsystem or supporting technology must be clearly described and its function in relationship to the greater EGS system must be expressed clearly. Approaches can be targeted at any of the multi-step project stages for technology development: from design concept, through scale model development (if applicable), to laboratory testing, field testing, and commercial scale demonstrations.

Grant applications are sought in the following subtopics:

a. High Temperature Downhole Logging and Monitoring Tools—Challenging subsurface conditions are one of the barriers to an accelerated ramp-up of geothermal energy generation. To address this challenge, grant applications are sought to develop logging and monitoring tools that are capable of tolerating extreme environments of high temperatures and pressures. The instruments of interest include, but are not limited to, temperature and pressure sensors, flow meters, fluid samplers, inclination and direction sensors, acoustic instruments (high and low frequency), resistivity probes, natural gamma ray detectors, epithermal neutron scattering gauges, rock density gauges (gamma and sonic), casing monitoring devices (e.g. cement bond logs and casing collar locators), fluid conductivity, pH indicators and well dimension probes (caliper). The target temperatures and pressures for these logging and monitoring tools should be supercritical conditions (374° C and 220 bar for pure water), and the tools may be used at depths of up to 10,000 meters.

Questions – Contact Raymond Fortuna, 202-586-1711, raymond.fortuna@ee.doe.gov.

b. Cements for EGS Applications—While conventional geothermal wells experience large temperature rises during production, EGS wells experience large temperature drops at the bottom of the well during the stimulation process, due to the cooling effect of the injected water. This temperature drop may be in the neighborhood of 350°F. This unique situation causes significant stress and potential failure of the cement sheath if conventional cement systems are utilized. To address this issue, grant applications are sought for the research, design, development, testing, and demonstration of a cement system for the high temperature and stress conditions of an EGS wellbore. Proposed approaches may define cement formulations that would be used by the geothermal industry to place the cement within a long string of casings; such approaches should focus on preventing a premature set and maintaining a strong seal at the shoe (so that stimulations may be performed through the casing).

Questions – Contact Raymond Fortuna, 202-586-1711, raymond.fortuna@ee.doe.gov.

c. Drilling Systems—High upfront costs, largely due to high drilling costs, are a major barrier to expanded geothermal energy production in the United States. Therefore, grant applications are sought to reduce drilling costs by developing a drilling technology (horizontal and/or directional) that is capable of drilling three times faster than conventional rotary drilling. Approaches of interest include, but are not limited to the design and development of improved drilling fluids (to reduce frictional viscosity and remove cuttings), high-performance bottom-hole assemblies (e.g., collars, bent subs, drill bits), and downhole motors (to control wellbore orientation). Proposed approaches must demonstrate reliable operation and equipment durability that exceeds the performance of conventional equipment at depths up to 10,000 meters and temperatures up to 300°C.

Questions – Contact Raymond Fortuna, 202-586-1711, raymond.fortuna@ee.doe.gov.

d. Fracture Characterization Technologies—Subsurface imaging is an important part of creating a productive EGS reservoir, which requires visualization before, during, and after creation. In order to advance technology and reduce the upfront risk to geothermal projects, more robust subsurface imaging technologies must be developed. Grant applications are sought to develop improved downhole and remote imaging methods to characterize fractures. Fracture characterization includes prediction of fracture and stress orientation prior to drilling (needed to properly orient horizontal wells); determination of fracture location, spacing, and orientation (while drilling); and determination of the location of open fractures (after stimulation), in order to identify the location of fluid flow pathways within the enhanced geothermal reservoir. Proposed approaches should address robust methods for interpreting and imaging the subsurface, including but not limited to, the development of active or passive seismic, processing software, and joint inversion of geophysical techniques.

Questions – Contact Raymond Fortuna, 202-586-1711, raymond.fortuna@ee.doe.gov.

e. Working Fluids for Binary Power Plants—Binary power plants are rapidly becoming a major part of the geothermal industry, due to increased development of lower temperature geothermal resources. To address cost barriers associated with the working fluids in these binary power plants, grant applications are sought to (1) identify non-azeotropic mixtures of working fluids for improved utilization of available energy in subcritical cycles; (2) characterize the composition and thermophysical and transport properties of those mixtures; (3) identify working fluids for supercritical cycles and trilateral cycles; and (4) characterize the composition, thermophysical, and transport properties of those working fluids. Proposed approaches may address working fluids or mixtures of working fluids with the potential for greater energy conversion efficiency than conventional working fluids, such as isobutane or refrigerants.

Questions – Contact Raymond Fortuna, 202-586-1711, raymond.fortuna@ee.doe.gov.

f. GHP Component R&D—High initial costs have been identified as a key barrier to widespread GHP deployment. To address this barrier, applications are sought to improve GHP components to increase efficiency as well as energy savings as compared to conventional systems. Applications may address but are not limited to: variable-speed (VS) components, advanced sensors and controls (including water flow sensing), electronic expansion valves, heat exchange (HX) design and fluids, system optimization, unit control algorithms, and load management tools.

Questions – Contact Raymond Fortuna, 202-586-1711, raymond.fortuna@ee.doe.gov

g. Innovative System/Loop Designs—One of the main barriers in GHP technology is the high cost of drilling and loop installation. Applications are sought for innovative system/loop designs that reduce the costs of system and/or loop installation, through new design layouts, system components, materials, and/or methods.

Questions – Contact Raymond Fortuna, 202-586-1711, raymond.fortuna@ee.doe.gov.

(b) Titles of Relevant CPC Classes

Technology Distance ptile.	Example CPC Titles
1	Geophysics; gravitational measurements Positive-displacement machines for liquids; pumps Collection, production or use of heat
10	Electric heating; electric lighting Static electricity; naturally-occurring electricity Cyclically operating valves for machines or engines
20	Installations or methods for obtaining, collecting, or distributing water Computer systems based on specific computational models Vehicles, vehicle fittings, or vehicle parts

Notes: Topic #4 from the FY2010 Release 1 Funding Opportunity Announcement.

(a) FOA Text

38. DATA MANAGEMENT AND STORAGE

a. Green Storage for HPC with Solid State Disk Technologies: From Caching to Metadata Servers—Most solid-state storage devices (SSDs) use non-volatile flash memory, which is made from silicon chips, instead of using spinning metal platters (as in hard disk drives) or streaming tape. By providing random access directly to data, the delays inherent in electro-mechanical drives are eliminated. The common consumer versions, known as flash drives, are compact and fairly rugged. Advantages attributed to SSDs include higher data transfer rates, smaller storage footprint, lower power and cooling requirements, faster I/O response times (up to 1000 times faster than mechanical drives), improved I/O operations per second (IOPS), and less wasted capacity.

Furthermore, upcoming processor chip designs from Intel and AMD will include SSD/FLASH controllers built on-board the CPU chip, in order to improve integration for laptop and embedded applications. Such technology is likely to enable a localized checkpoint-restart capability to mitigate increased transient failure rates on future ultra-scale computing systems. This increased level of hardware integration makes it clear that x86 server nodes, which incorporate SSD directly onto the node, are on the horizon.

In view of these developments, the DOE seeks to improve its understanding of the implications of SSDs for large-scale, tightly-coupled systems in High Performance Computing (HPC) environments. Therefore, grant applications are sought to further develop SSD technology as a cost-effective and productive storage solution for future HPC systems, including, but not limited to:

1) **Categorization of SSD failure modes** - The rate of deployment of SSDs in HPC environments will be artificially slowed until a better understanding of the failure modes of this new class of storage is achieved. Proposed approaches should categorize the type of failure (wire bond, cell wear-out, or other failure) and determine how the failures would be detected and/or repaired in a composite device fielded in an HPC environment.

2) **Use of SSD for node-local storage, for faster (localized) checkpoint/restart (CPR)** - If transient failures cause nodes to die, then SSD could be a viable approach for fault-resilience. However, for nodes subjected to hard-failures, the use of SSD could produce an even higher node failure rate, due to the inherent failure characteristics of the SSD; in this case, the SSD approach would not be viable for CPR. Approaches of interest should collect and analyze data on the known failure modes of existing SSD components vis-a-vis node failure modes, in order to determine if SSD presents an effective alternative to the checkpoint/restart of a shared file system.

3) **Use of SSD for scalable out-of-core applications** - Although node-local disk systems have been used to support some applications that use out-of-core algorithms (such as some components of NWChem), the failure rates of spinning disks have rendered this practice unfeasible. Rather, central file systems are used to support these out-of-core applications, greatly affecting their scalability. Approaches are sought to determine whether local SSD might be reliable enough to enable a scalable approach to out-of-core processing.

4) **Use of SSD for metadata servers** - Metadata servers subject disk subsystems to many very small transactions, a feature that is very difficult to support with existing mechanical/spinning-disk based systems. SSDs might respond better to the random-access patterns required for metadata servers, but may not perform as well for write functions. Approaches of interest should analyze the data access patterns of a typical HPC Lustre metadata server and, using an SSD performance model, determine how well an SSD-based system would respond to a metadata server load.

5) **Use of SSD for accelerated caching for the front-end of large-scale disk arrays** – The use of SSDs in caching for large-scale disk arrays is an emerging technology that is not well understood. Approaches are sought to determine of both its performance potential when subjected to real workloads and its fault resilience.

b. Data Management Tools for Automatically Generating I/O Libraries—Database-like, self-describing, portable binary file formats, such as Network Command Data Form (NetCDF) and Hierarchical Data Format (HDF), greatly enhance scientific I/O systems by raising the level of abstraction for data storage to very high-level semantics (of data schemas and relationships between data objects stored) rather than low-level details of the location of each byte of the data stored in the file. However, both NetCDF and HDF5 still rely on very complex APIs to describe the data schema, and many performance pitfalls can arise if the APIs are not used in an optimal manner. Consequently, application developers must invest considerable effort in creating their own "shim" I/O APIs that are specific to their applications, in order to hide the complexity of the general-purpose APIs of NetCDF and HDF5.

Grant applications are sought to develop software tools that not only would enable rapid prototyping of high-level data schemas but also would automatically generate a high-level API for presentation to application developers, thereby hiding the complexity of the low-level NetCDF and HDF5 APIs for managing the file format. Such tools also might use auto-tuning techniques to find the best performing implementation of an I/O method.

c. Integration of Scientific File Representations with Object Database Management Systems—Scientific file formats like Network Command Data Form (NetCDF) and Hierarchical Data Format (HDF5) have capabilities that closely match those of commercial Object Database Management Systems (ODBMS); yet, commercial ODBMSs provide much more sophisticated data management tools than are available to users of NetCDF and HDF5. Unfortunately, ODBMSs are not designed to accommodate parallel writes to the same data entry from multiple parallel writers. Furthermore, database storage formats are opaque and non-portable, and no file standard exists to facilitate the movement of data from one database system to another. By contrast, NetCDF and HDF5 both offer open, standardized formats and portable, self-describing binary formats for storing data represented as Object Databases.

(b) Titles of Relevant CPC Classes

Technology Distance ptile.	Example CPC Titles
1	Electric digital data processing Apparatus or arrangements for taking photographs or for projecting or viewing them Transmission of digital information, e.g. telegraphic communication
10	Information and communication technology adapted for specific application fields Radio-controlled time-pieces Secret communication; jamming of communication
20	Presses in general Production of cellulose by removing non-cellulose substances Methods of steam generation; steam boilers

Notes: Topic #38 from the FY2010 Release 1 Funding Opportunity Announcement.

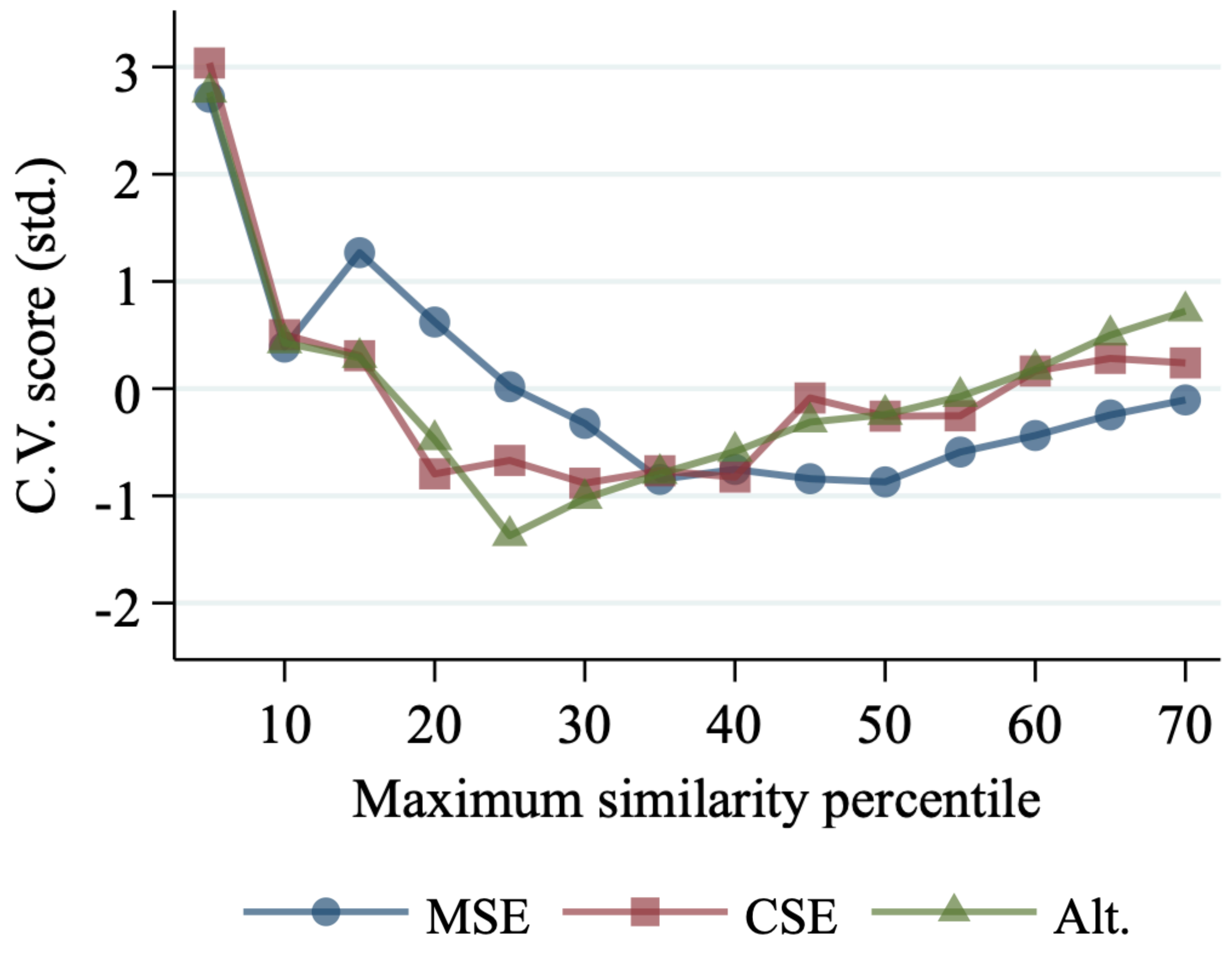
Mapping investments (SBIR $) to technology-space (CPC codes)

face validity of text-similarity mapping

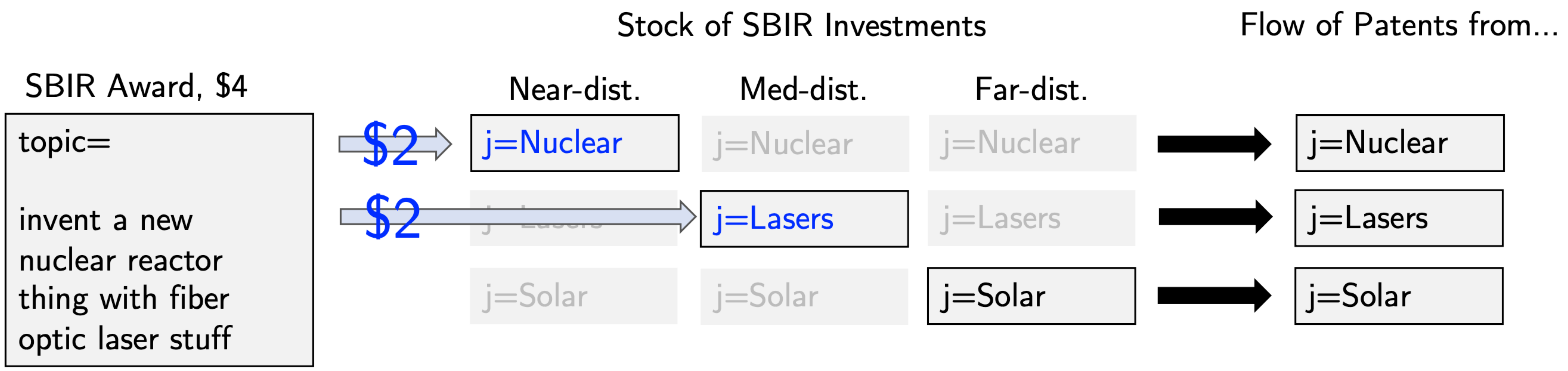
Funding Rank	CPC 3-digit Title
1	G01: measuring; testing
2	H01: basic electric elements
3	H02: generation; conversion or distribution of electric power
4	H03: basic electronic circuitry
5	H04: electric communication technique
6	G06: computing; calculating; counting
7	C10: petroleum, gas or coke industries; technical gases ...
8	F16: engineering elements and units...
9	C12: biochemistry; microbiology; enzymology...
10	B60: vehicles in general
11	F02: combustion engines; hot-gas engine plants
12	B01: physical or chemical processes or apparatus

Empirical model
determining boundaries of spillovers

- **Iterate**:
  1. Assume spillovers stop after ___ distance
  2. Estimate model
  3. Recover goodness-of-fit
  4. Repeat (1-3), and pray for “convergence”



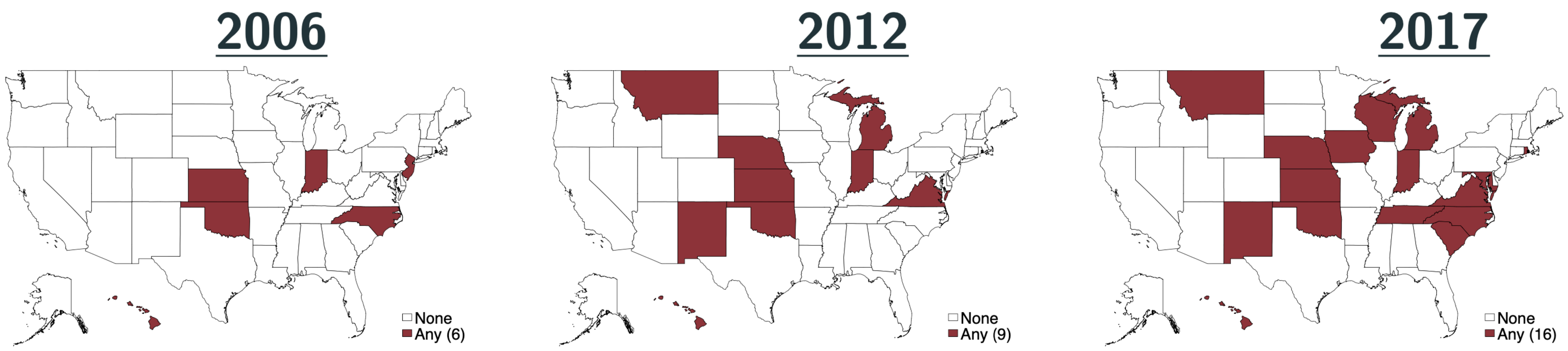
Empirical model
patent production function



Empirical model

exogenous investments in space-years

- **State-specific match programs**: if located in state with match, recipient firm receives a "bonus" valued at 25-100% of the federal SBIR award
- **Key assumption**: firms (and the tech. they're pursuing) in match policy states are not more/less productive than avg.



Empirical model

identifying technological and geographic spillovers

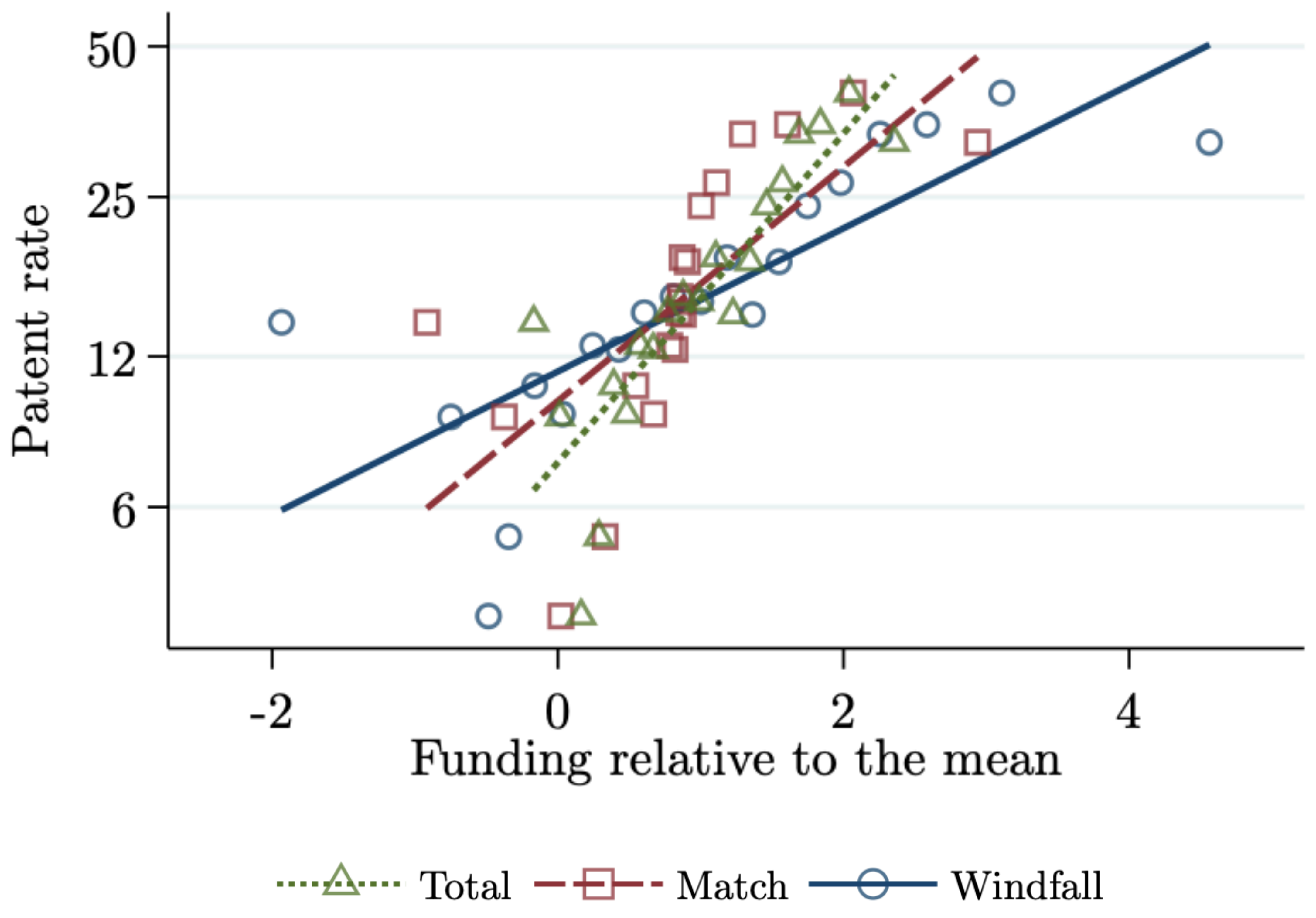
$$\mathbf{E}[\, y_{jt}^{d} \mid W_{jtb} \,] = \exp\left(\sum_{b \in B} W_{jtb} \theta_b^d + \tau_t^d \right)$$

- **Technological spillovers**:
  - Count only output from the set of producers who are distance d from SBIR grant recipients
  - ***Within a single regression***, see how space-time level output depends on how similar (per b) the investments where in space-time — compare θ_b parameters in the same regression: $\theta_{b=more\ sim\ tech}^{d=nearby\ firms}$ vs. $\theta_{b=less\ sim\ tech}^{d=nearby\ firms}$
- ***Geographic spillovers***:
  - Focus on investment-output relationship of some fixed amount of similarity (per b)
  - ***Across regressions***, see how output depends on which producers' output is included — compare θ_b parameters in the same regression: $\theta_{b=more\ sim\ tech}^{d=nearby\ firms}$ vs. $\theta_{b=more\ sim\ tech}^{d=distant\ firms}$

Results: Evidence of endogenous funding

binscatters of investment stocks and patent flows

Figure E.1: Patenting and Funding Conditional on Aggregate Time Trends



Results

spillovers are large: productivity depend on what “counts”

Table 3—Summary of Outputs and Costs

	% of net patents	$\frac{\text{patents}}{\$1\text{M}}$	$\frac{\$}{\text{patent}}$
Counting all USPTO patents and			
… only grant recipients	26	0.75	$ 1,330,000
… only nonrecipient firms and inventors nearby recipients	20	0.59	$ 1,684,000
… only remainder of US nonrecipients	14	0.40	$ 2,476,000
… all US firms and inventors	59	1.75	$ 572,000
… all foreign firms and inventors	41	1.19	$ 839,000
Counting all firms and inventors, only USPTO patents that are			
… very similar to grants’ tech. objectives	37	1.10	$ 909,000
… somewhat similar to grants’ tech. objectives	40	1.17	$ 853,000
… least similar to grants’ tech. objectives	23	0.67	$ 1,496,000
Counting all USPTO patents, all firms and inventors	100	2.94	$ 340,000

Notes: Reports average marginal products and costs when focusing on a particular set of patents or firms and inventors. The bottom row defines output and costs when all patents are considered, so “% of net patents” is 100 percent by construction; “patents/$1M” reports the net number of patents expected from a marginal investment (awarded only to grant recipients) of $1 million; “$/patent” reports the marginal cost expected to produce one additional patent.

Additional result: “Value”

a slightly closer look at externality (but still var from **externality**)

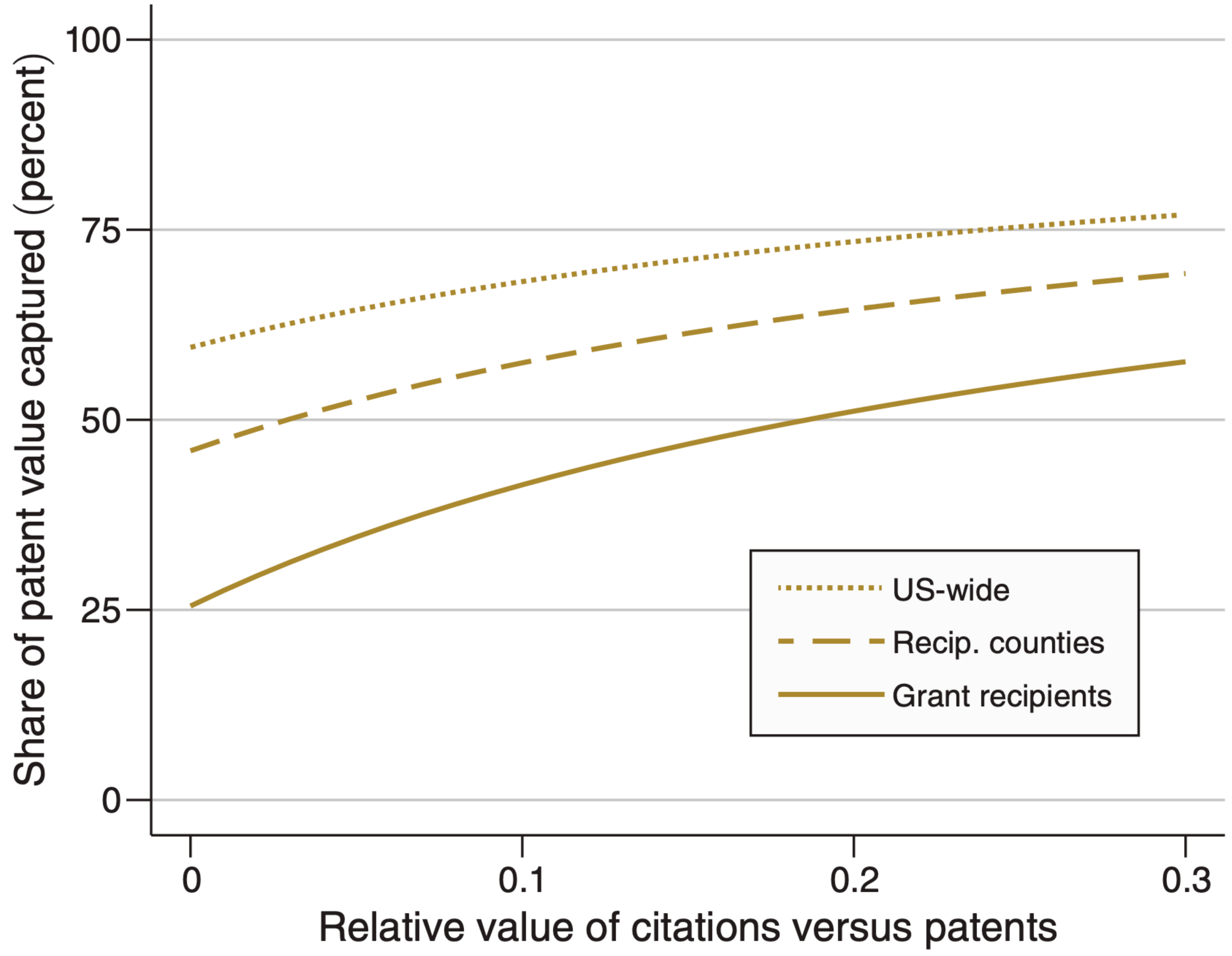
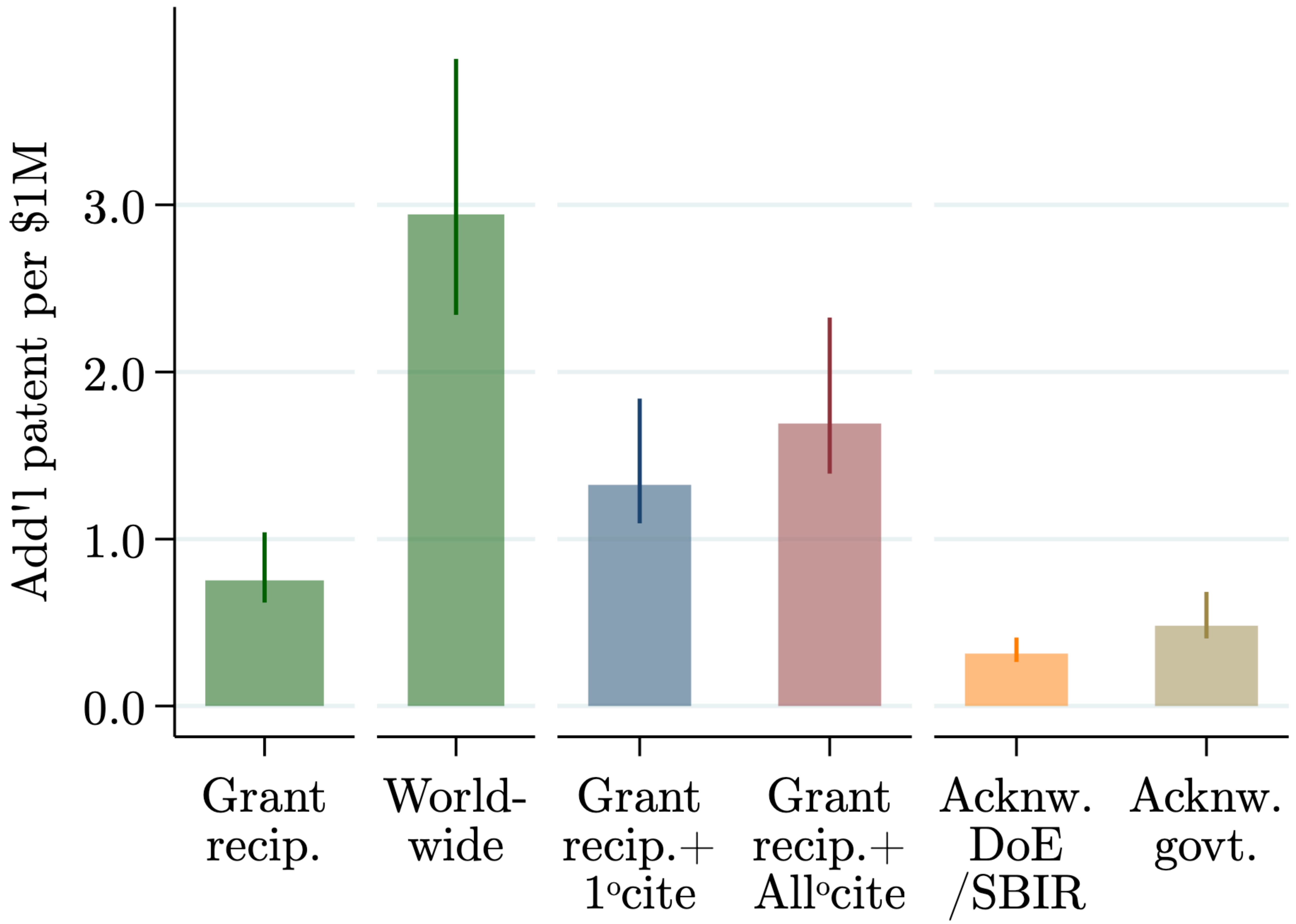


FIGURE 4. SHARE OF NET PATENT VALUE CAPTURED BY DIFFERENT FIRMS AND INVENTORS

Additional result: Identifying spillovers

paper trails are very misleading



But!

some (very difficult) unanswered questions

- **“Externalities” versus “Spillovers”**
  - How much value do scientists appropriate?
  - How does this vary ex-ante (at time of investment) vs. ex-post (after discovery)?
- **Dynamics**
  - What is the time between investment and payoff and what determines this?
- **Heterogeneity**
  - What are the specific, economic fundamentals of technologies that lead to larger/smaller externalities and/or spillovers?

Scientists as choosers

preferences and adjustment costs

- **Stern**. "Do scientists pay to be scientists?"

  Management Science 50, no. 6 (2004): 835-853.

- **Myers**. "The elasticity of science.”

  American Economic Journal: Applied Economics 12, no. 4 (2020): 103-134.

- **Acemoglu**. "Diversity and technological progress.”

  The Rate and Direction of Inventive Activity Revisited (2011). U. Chicago Press, 319-356.

Aside:
Estimating Demand in Science

We are often focused on scientists' (demand) choices

- And these choices can often be formulated as a discrete choice problem
  - What science to study?
  - What collaborator to work with?
  - What journal to submit to?
  - What results to report? [note: continuous things here too; e.g., *p-hacking*)
- Estimate or motivate (or both)
  - e.g., Krieger, Myers, & Stern. *"How Important is Editorial Gatekeeping? Evidence from Top Biomedical Journals" Review of Economics and Statistics (forthcoming).*

Often, scientists' "demand" = "entry"

- Standard IO market entry model
  - Decision-maker: **scientists**
  - Competition: **other scientists**
  - Market: **geographic location; science space**
  - Market features: **consumer demand; fixed & variable costs of entry**
- See:
  - **Seim**. *"An empirical model of firm entry with endogenous product-type choices." The RAND Journal of Economics (2006).*
  - **Bajari, Hong, Krainer, & Nekipelov**. *"Estimating static models of strategic interactions. Journal of Business & Economic Statistics (2010).*

Do Scientists Pay to be Scientists?

Scott Stern

Management Science 50, no. 6 (2004): 835-853.

Compensating differentials
why are they important?

- **Earnings inequality**

⇒ labor market policies

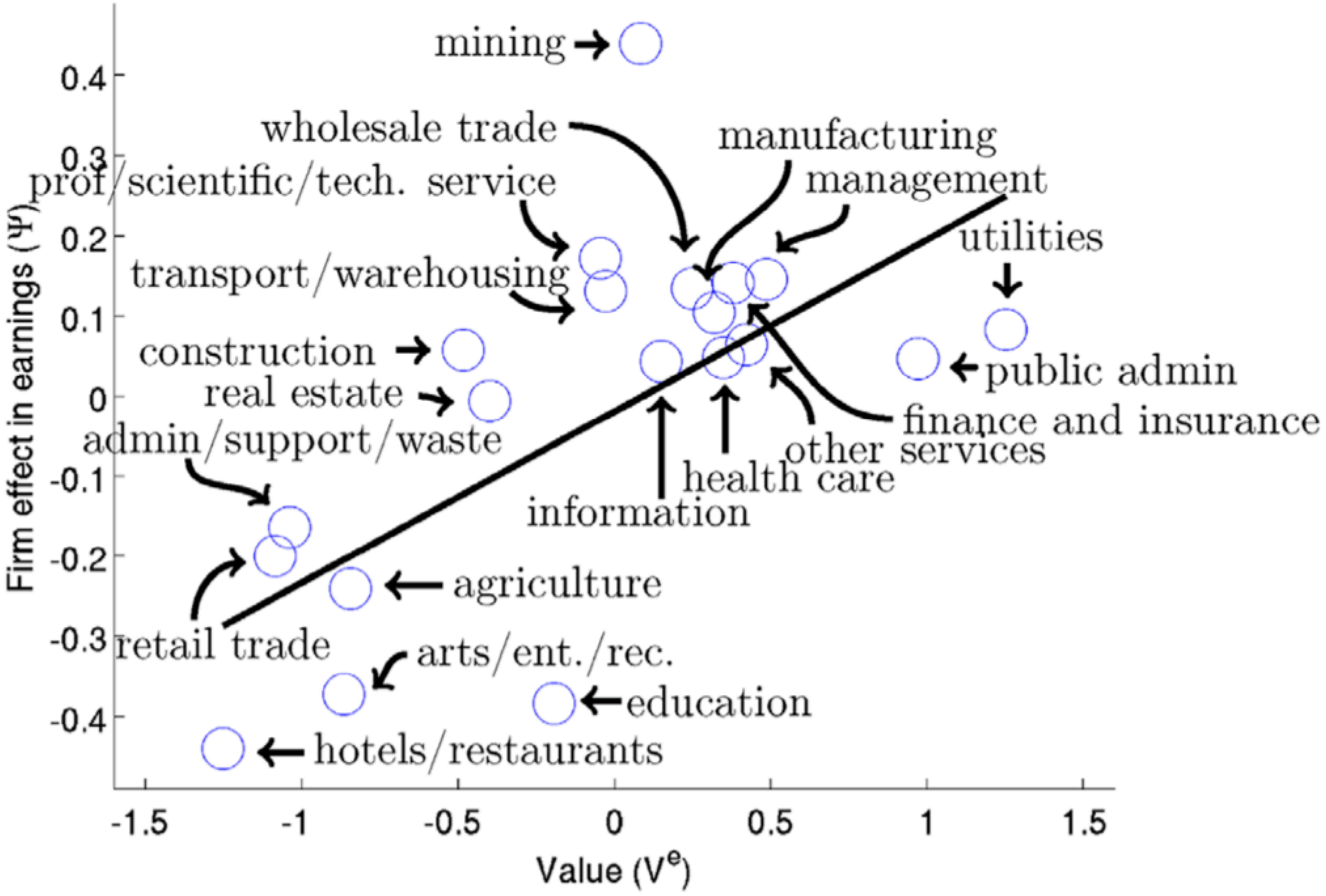


FIGURE V

Relationship between Values and Earnings

Sorkin, Isaac. "Ranking firms using revealed preference." (2018).

Compensating differentials
why are they important?

- **Earnings inequality**

⇒ labor market policies

- **Contract design**

⇒ incentives for innovation

RAND Journal of Economics
Vol. 39, No. 3, Autumn 2008
pp. 617–635

Academic freedom, private-sector focus, and the process of innovation

Philippe Aghion*
Mathias Dewatripont**
and
Jeremy C. Stein***

We develop a model that clarifies the respective advantages and disadvantages of academic and private-sector research. Rather than relying on lack of appropriability or spillovers to generate a rationale for academic research, we emphasize control-rights considerations, and argue that the fundamental tradeoff between academia and the private sector is one of creative control versus focus. By serving as a precommitment mechanism that allows scientists to freely pursue their own interests, academia can be indispensable for early-stage research. At the same time, the private sector's ability to direct scientists toward higher-payoff activities makes it more attractive for later-stage research.

Stern (2004): The model
scientists' utility and firm profits from a job

Scientist utility: $U_{ij} = \alpha \gamma_i \mathbf{1}\{\text{science}_{ij}\} + w_{ij}$

Firm profits: $\pi_{ij} = \beta \gamma_i \mathbf{1}\{\text{science}_{ij}\} - w_{ij} - \delta \mathbf{1}\{\text{science}_{ij}\}$

- Scientist's taste for science: α
- Firm's revenues from science: β
- Scientist's ability: γ_i
- Job's scientific orientation: science_{ij} , 1=yes, 0=no
- Wage: w_{ij}
- Firm's cost of science: δ

Stern (2004): The model

equilibrium wages

$$w_{ij}^* = \gamma_i + \gamma_i (\phi \beta - \alpha) \mathbf{1}\{\text{science}_{ij}\}$$

- Scientist's taste for science: α
- Firm's revenues from science: β
- Scientist's ability: γ_i
- Job's scientific orientation: science_{ij} , 1=yes, 0=no
- Wage: w_{ij}
- Rent-splitting parameter (share going to scientists): $\phi \in (0,1)$

Stern (2004): The model

firm's decision to do science

$$\mathbf{1}\{\text{science}_{ij}\} = 1 \;\; iff \;\; \gamma_i > \frac{\delta}{(1-\phi)(\beta-\alpha)}$$

- Firms offer more science when, ceteris paribus:
  - **scientists are higher quality**
  - when cost of science (δ) is low
  - when share of quasi-rents captured by scientists (ϕ) is low
  - when revenue from science (β) is high
  - when taste for science (α) is low

The model and regression

equilibrium wages

$$w_{ij}^* = \gamma_i + \gamma_i (\phi\beta - \alpha)\,\mathbf{1}\{\text{science}_{ij}\}$$

Regression: $w_{ij}^* = \theta_0 + \theta_{\text{s}}\mathbf{1}\{\text{science}_{ij}\} + \epsilon_{ij}$

Problem: $\mathbf{1}\{\text{science}_{ij}\}$ may be corr. with γ_i which is corr. with w_{ij}^*

$\Rightarrow \mathbf{1}\{\text{science}_{ij}\}$ may be corr. with ϵ_{ij}

The model and regression

equilibrium wages

$$
\begin{aligned}
w^*_{ij} &= \gamma_i + \gamma_i (\phi\, \beta - \alpha)\, \mathbf{1}\{\text{science}_{ij}\} \\
&= \gamma_i \left(1 + (\phi\, \beta - \alpha)\, \mathbf{1}\{\text{science}_{ij}\}\right)
\end{aligned}
$$

Regression w/ FE: $w^*_{ij} = \theta_i + \theta_{\text{S}} \mathbf{1}\{\text{science}_{ij}\} + \epsilon_{ij}$

$$\theta_{\text{S}} \propto (\phi\, \beta - \alpha)$$

Key assumptions
behind connection between empirical model

- Observed offers are equally “serious” [tried to get “final round” offers]
- Multiple-offer scientists are representative of single-offer scientists [Table 6A]
- Firms have equal view of scientists’ quality γ_i [survey design]

Recall: $\mathbf{1}\{\text{science}_{ij}\} = 1 \;\; iff \;\; \gamma_i > \dfrac{\delta}{(1-\phi)(\beta-\alpha)}$

Key: Conditional on γ_i, variation in scientific orientation of offer is driven by…

Stern (2004): The results

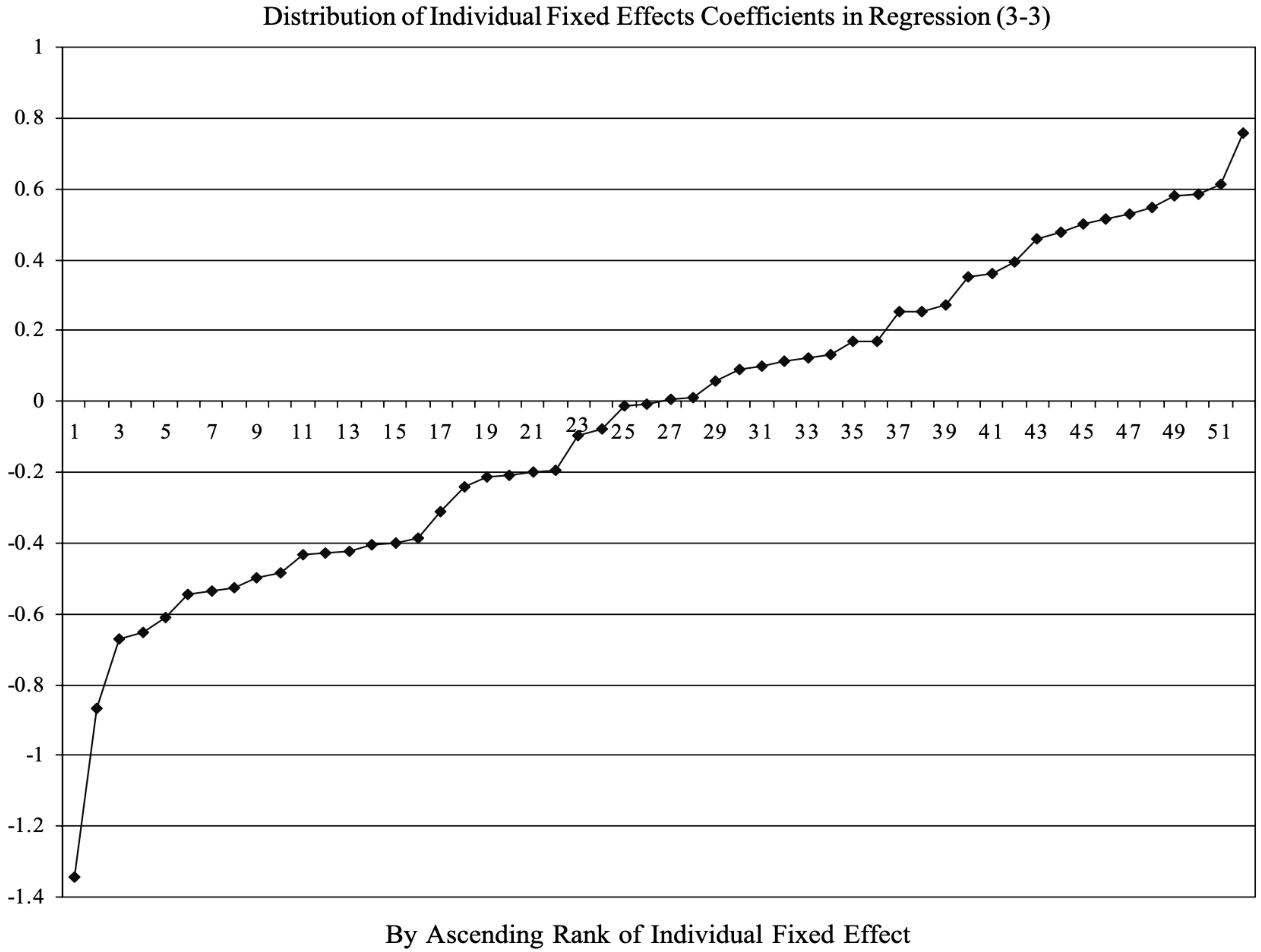
scientists' salary offer given job features

Table 3 Hedonic Wage Regression: Overall Sample Dependent Variable = LN(SALARY), # of Observations = 121

	Permission to publish			Combination model	Science index model	
	(3–1)	(3–2)	(3–3)	(3–4)	(3–5)	(3–6)
	Baseline (NO FE)	Baseline (w/FE)	Full model (w/FE)	Full model (w/FE)	Full Model (w/FE)	Full Model (w/FE)
PERMIT_PUB	0.027 (0.186)	−**0.266** (**0.114**)	−**0.191** (**0.105**)	−0.089 (0.103)		
CONTINUE RESEARCH				−**0.134** (**0.060**)		
INCENT_PUB				−0.036 (0.028)		
SCIENCE INDEX					−**0.114** (**0.053**)	−0.078 (0.057)
EQUIPMENT				**0.063** (**0.033**)	**0.057** (**0.030**)	**0.053** (**0.031**)
JOBTYPE CONTROLS	no	no	yes (5; Sig.)	no	no	yes (5)
Individual fixed effects	no	yes (52; Sig.)	yes (52; Sig.)	yes (52; Sig.)	yes (52; Sig.)	yes (52; Sig.)
R-squared	0.001	0.915	0.955	0.958	0.954	0.958

Notes. Only persons with multiple job offers are included.
Standard errors are shown in parenthesis; significant coefficients (10%) are shown in bold.
Sig. stands for joint significance of fixed effects or job type controls (at 10% level).

Stern (2004): The results
distribution of scientists' fixed effects



The Elasticity of Science

Kyle Myers

American Economic Journal: Applied Economics 12, no. 4 (2020): 103-134.

Motivation: The Elasticity of Science
(ex-post rationalization of PhD madness)

- An economy is (generally) more efficient when producers face low adjustment costs
  - Demand shifts → the fast supply catches up, the better
- e.g., the clean energy transition
  - *Acemoglu, Aghion, Bursztyn, & Hemous. "The environment and directed technical change." American Economic Review (2012).*
  - *Aghion, Dechezleprêtre, Hemous, & Van Reenen. "Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry." Journal of Political Economy (2016).*

What actually happened ...

Requests For Applications (RFAs)

an example

Title: **Development of New Technologies Needed for Studying the Human Microbiome (R01)**

Announcement Type
This Funding Opportunity Announcement (FOA) is a reissue of RFA-RM-08-026.

Request for Applications (RFA) Number: **RFA-RM-09-008**

Key Dates
Release/Posted Date: July 16, 2009
Opening Date: August 14, 2009 (Earliest date an application may be submitted to Grants.gov)
Letters of Intent Receipt Date(s): August 17, 2009
NOTE: On-time submission requires that applications be successfully submitted to Grants.gov no later than 5:00 p.m. local time (of the applicant institution/organization).
Application Due Date(s): September 14, 2009
Peer Review Date(s): February-March 2010
Council Review Date(s): May 2010
Earliest Anticipated Start Date(s): July 2010
Additional Information To Be Available Date (Activation Date): Not Applicable
Expiration Date: September 15, 2009

Requests For Applications (RFAs)

an example

Executive Summary

- **Purpose.** The purpose of this FOA is to solicit applications to develop new and improved technologies for obtaining samples of individual microbial isolates or strains, from the human microbiota, suitable for complete genomic sequence analysis. The goal is to expand the number of “reference” microbial genome sequences, which in turn will aid in the analysis of the complex microbial populations resident in and on the human body.
- **Mechanism of Support.** This FOA will utilize the NIH Research Project Grant (R01) grant mechanism and runs in parallel with a FOA of identical scientific scope, RFA-RM-09-009 that solicits applications under the R21 mechanism.
- **Funds Available and Anticipated Number of Awards.** $2 million is available in FY10 for this FOA and the parallel R21 FOA in combination. It is anticipated that 2-4 R01 grants (of duration up to 3 years) and 2-6 R21 grants will be awarded. . Awards issued under this FOA are contingent upon the availability of funds and the submission of a sufficient number of meritorious applications.
- **Budget and Project Period.** Because the nature and scope of the proposed research will vary from application to application, it is anticipated that the size and duration of each award will also vary. Applicants for R01 grants may request a project period of up to 3 years.

Requests For Applications (RFAs)

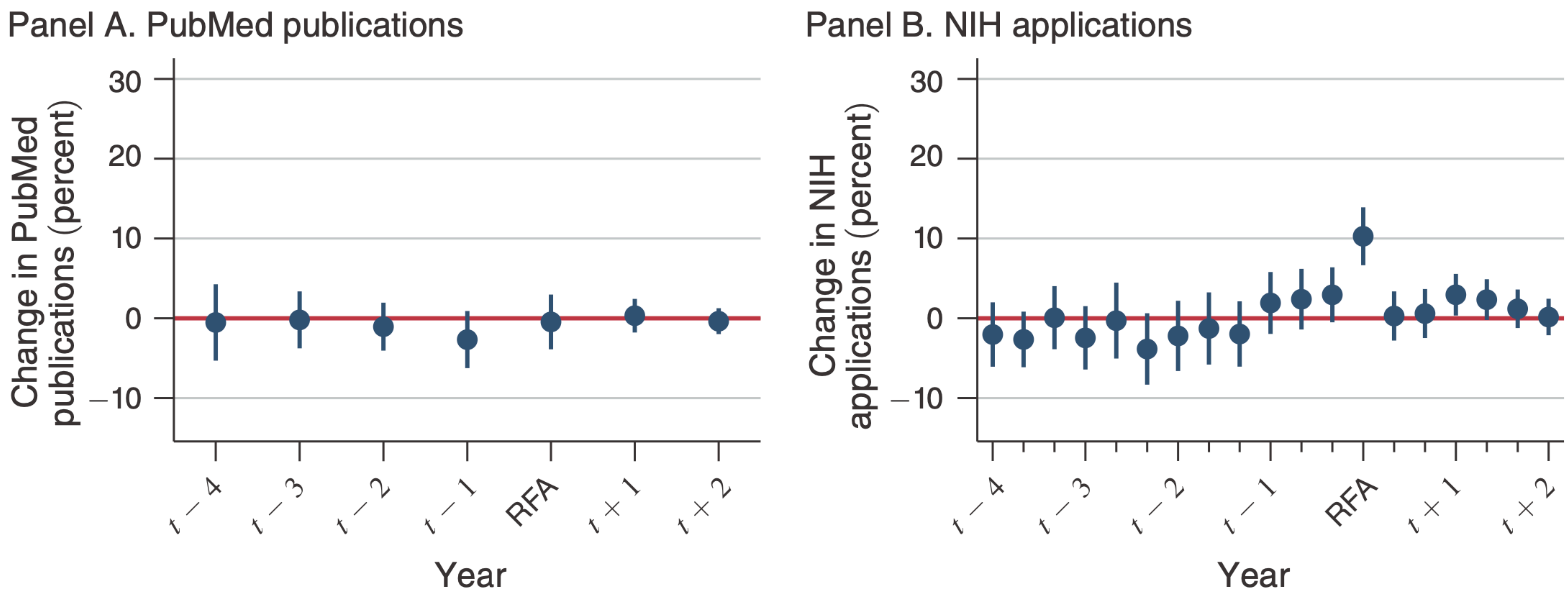
an example

RESEARCH SCOPE: The interpretation of metagenomic sequence data is greatly aided by comparison to the genomic sequence of isolated species and genetically different strains of the same species. Yet, only a small proportion of the microbial species resident in or on the human body has been isolated and sequenced. The purpose of this FOA is to support the development of technologies that will allow the determination of the complete, individual genome sequences of substantial numbers of previously uncharacterized members of the human microbiota, to aid in the interpretation of metagenomic datasets obtained from sampling the human body. The following list, which is certainly incomplete, presents examples of strategies that would be supportable under this FOA:

- Development of methods to isolate single microbial cells. These methods would enable the identification, analysis and isolation of individual cells in the human microbiota that satisfy a specified set of criteria.
- New approaches to obtain pure cultures or simple mixed cultures of small numbers of previously uncultivated species would advance the objective of genomic analysis of the human microbiota. Proposed methods that can be applied to a large number of species rather than to any one particular species will take high priority.
- Development, optimization and validation of methods to isolate, amplify, or clone unamplified or amplified DNA of whole genomes from individual cells at high fidelity (e.g., complete coverage, low bias, low chimerism).
- Development of methods to "normalize" the complexity of the population, at either the cellular or DNA level. Such methods would facilitate either the ability to isolate single cells that are rare within a population, or to perform bioinformatics analysis on metagenomic sequences (e.g., by improving the representation of "rare" members).
- Development of methods to enrich the cells of a given species to essential purity. This is the inverse of reducing redundancy, and might be most effective for species whose abundance is already high. Such methods might substitute, at least for DNA sequencing studies, for the ability to establish pure cultures.
- Development of methods that (as a prelude to isolating single microbial cells, or conducting enrichment or normalization) disaggregate cells from the complex mixtures of microbial cells, human cells, and extracellular materials (e.g., biofilms) that comprise human microbial samples. Methods for cell disaggregation should be developed in conjunction with associated methods such as those described above.

RFAs don’t appear to target “hot” topics

regression results



Scientists like being “close” to “big” RFAs

raw data

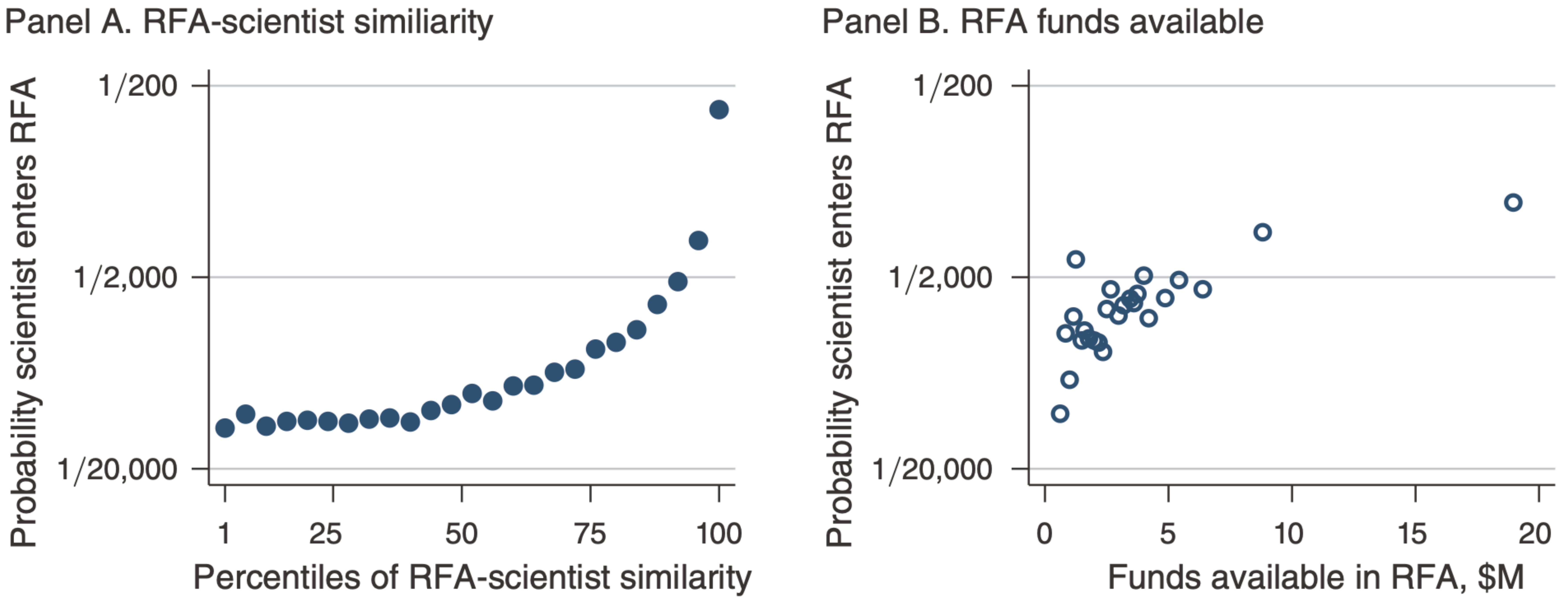
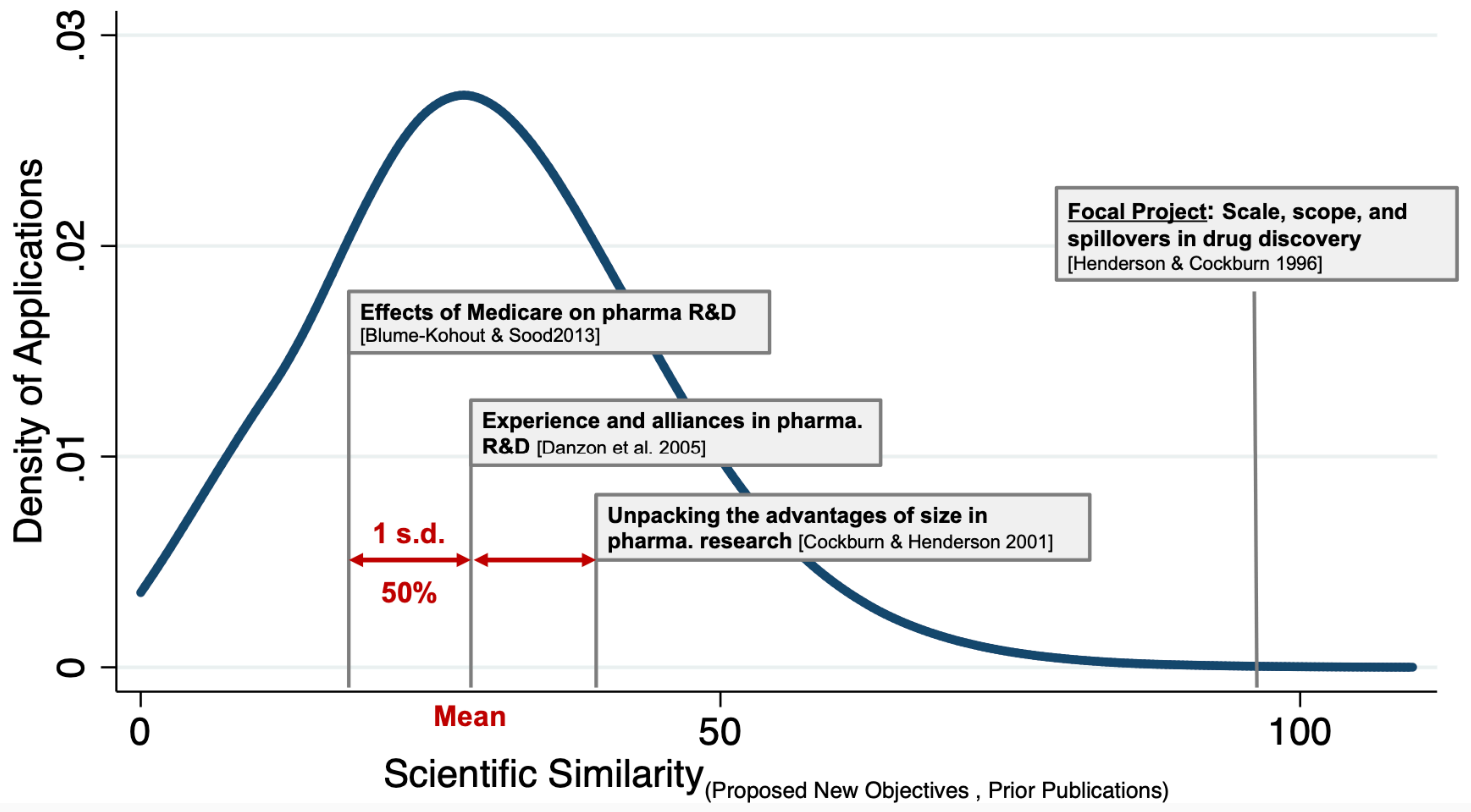


FIGURE 1. PROBABILITY OF RFA ENTRY PER SIMILARITY AND FUNDING

Notes: The figure shows binned scatterplots of entry probabilities per panel A, similarity of scientists’ prior publications to the research objectives of the RFA (larger scores indicate greater overlap), and panel B, the amount of funds made available in the RFA. The figure is based on approximately 110,000 scientists and 390 RFAs. Note the log scale of the *y*-axis.

Measuring Scientific Similarity
(and communicating it too)

Figure III.6: *pmra* Distribution: Economics Examples

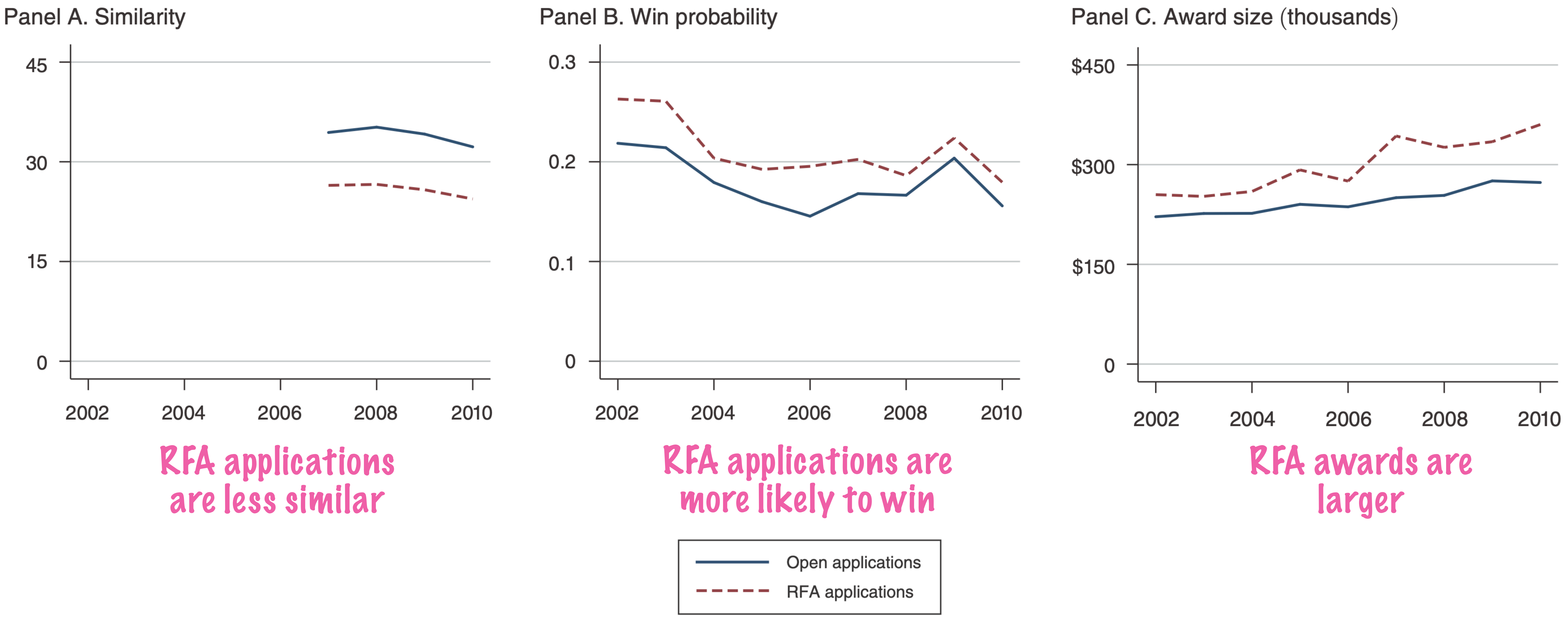


Adjustment Costs could be Large!
but, are they policy-relevant?

- Two major channels at the NIH:
  - “Investigator-initiated” / “open”: propose (almost) whatever you want
  - RFAs: propose something within the scope of objectives
- If adjustment costs are first-order and there aren’t a ton of scientists close to each RFA, then in equilibrium:
  - Scientists will see the RFAs and compare the extra adjustment costs relative to the extra expected payoff
  - But, they will never fully dissipate all (expected) rents in the RFAs
  - And, the size of those rents will equal the adjustment costs

Expected Costs and Benefits

RFAs versus Open channels



A Simple Entry Model to Estimate Adjustment Costs
handling competitive expectations

- **Concern**: If scientists like RFAs that are bigger ($) …
  - … ***scientists will know*** that RFAs with larger "purses" will attract many others …
    - … which increases competitive expectations …
      - … which could mute the effect of purse size on Pr(apply)
- **Concern**: If scientists like RFAs that are (scientifically) similar…
  - … ***scientists will know*** that RFAs in dense areas will attract many others …
    - … which increases competitive expectations …
      - … which could mute the effect of scientific similarity on Pr(apply)

A Simple Entry Model

handling competitive expectations (Bajari et al. 2010)

- Estimate scientists' expectations of how many others will enter:
  - $\Pr(Entry_{ij}) = a + \beta Similarity_{ij} + \gamma Purse_j + \epsilon_{ij}$
  - $\mathbf{E}[\Pr(Entry_{ij})] = \widehat{a} + \widehat{\beta} Similarity_{ij} + \widehat{\gamma} Purse_j$
  - $\widetilde{n}_{ij} = \sum_{i' \neq i} \left(\widehat{a} + \widehat{\beta} Similarity_{i'j} + \widehat{\gamma} Purse_j \right)$
- Estimate scientists' own probability of entering, given these expectations:
  - $\Pr(Entry_{ij}) = \alpha + \sigma Similarity_{ij} + \phi Purse_j + \delta \widetilde{n}_{ij} + \varepsilon_{ij}$

Results: Entry Model

TABLE 1—DETERMINANTS OF RFA ENTRY

	$\mathbf{1}\{Entry_{ij}\}$				
	(1)	(2)	(3)	(4)	(5)
$Purse_j$				$\boxed{2.32}$	$\boxed{4.07}$
				$\boxed{(0.551)}$	$\boxed{(0.503)}$
$Similarity_{ij}$				2.33	2.55
				(0.911)	(0.964)
$Competitive\ Expectations_{ij}$					−4.37
					(0.271)
Includes similarity bins					
RFA controls				Y	Y
Scientist fixed effects				Y	Y

ignoring competition severely biases (downward) responsiveness to $

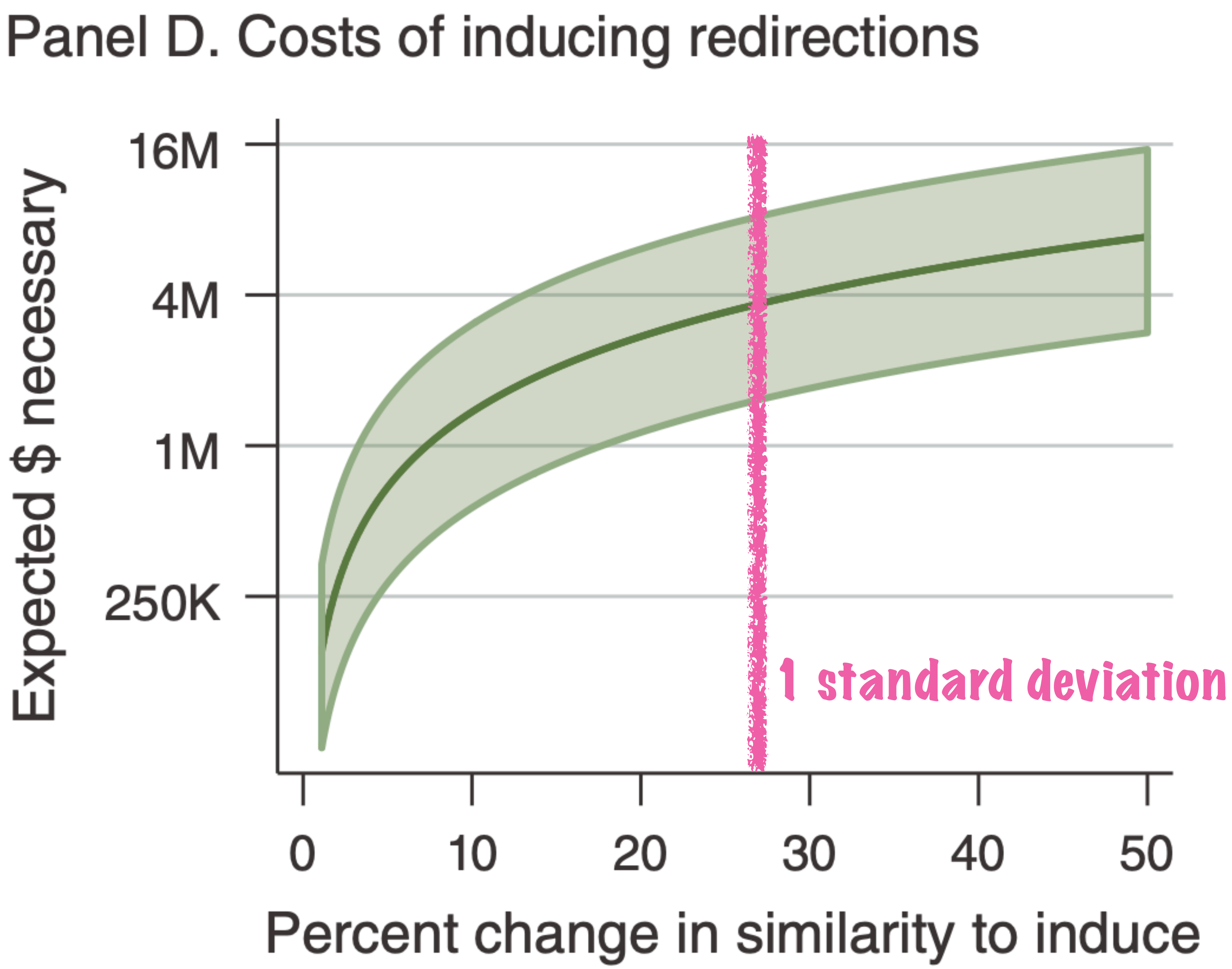
Notes: All models include 20,221,541 scientist-RFA (ij) pair observations, where the mean entry probability is 5.47×10^{-4}. Independent variables are standardized in regression, so coefficients indicate the change in entry probability associated with a one standard deviation increase in the variable; all coefficients are scaled by 10^{-4}.

The Elasticity of Science
from entry model parameters to adjustment costs

- Estimate scientists' own probability of entering, given these expectations:
  - $\Pr(Entry_{ij}) = \alpha + \sigma Similarity_{ij} + \phi Purse_j + \delta \widetilde{n}_{ij} + \varepsilon_{ij}$
  - $\sigma \equiv \partial \Pr(Entry)/\partial Similarity$
  - $\phi \equiv \partial \Pr(Entry)/\partial Purse$
- Elasticity of science: the percent change in scientific similarity that can be induced with a percent change in (expected) funding
  - EoS: $\dfrac{\sigma/S}{\phi/P}$

How much $ does it take?

elasticity of science ≈ 0.1



Are re-directions persistent?

yes

TABLE 4—GRANT PRODUCTIVITY—PUBLICATION SIMILARITY

	IHS(*Publication-RFA Similarity*$_{jk}$)		
	(1)	(2)	(3)
$\mathbf{1}\{Win, RFA_{jk}\}$	0.131	0.334	0.317
	(0.0328)	(0.166)	(0.136)
Semielasticity *RFA*	0.140	0.378	0.361
Observations	4,949	4,949	4,949
IV		Y	Y
F-statistic		57.5	58.2
Project, people **X**			Y
Funding group fixed effects	Y	Y	Y
pmra controls	Y	Y	Y
LASSO $var_{sel/poss}$	3/21	6/21	12/350

Summary & Take-aways

Myers (2020). *"The Elasticity of Science"*

- **The adjustment costs of modern (biomedical-like) science are very large**
  - In both absolute terms, and relative to current grant sizes
- Targeted funding mechanisms:
  - Give rents to scientists who apply
  - Cause significant changes in trajectory for winners
  - Cause as many (if not more) total publications compared to "open" channels
  - ⇒ **there could be a pseudo-deadweight-loss of intervening in science with $**
    - *[caveat: on the scale of how RFAs are used at the NIH in this period]*
    - *[caveat: don't forget Sampat (2012) Hegde & Sampat (2015)]*

Diversity and Technological Progress

Daron Acemoglu

The Rate and Direction of Inventive Activity Revisited (2011)

Simple Model: Setup
Acemoglu (2011)

- Two periods $t = \{1,2\}$; no discounting
- Two technologies j (sellable at $t = 1$) and j' (un-sellable at $t = 1$)
  - Sellable ("active"): if scientist makes improvement, they're rewarded
  - At $t = 1$, "quality" of both technologies = 1
- A scientist as 1 unit of time, can devote some share x to studying tech.
  - Quality of tech. improves with prob. $h(x)$; $h()$ is concave and well-behaved
  - Improvement moves quality from 1 to $(1 + \lambda)$, where $\lambda > 0$
  - Receive payoff of $(1 + \lambda)$ if successful

Expected payoff

x_j : scientists share of effort devoted to tech. j (note: $x_{j'} = 1 - x_j$)

v : prob. other scientist wins in either tech.

p : prob. of switch from tech. j to j'

$$\pi(x_j) = \underbrace{h(x_j)}_{\text{prob. win } j} \times \underbrace{(1+\lambda)}_{\text{payoff in } t=1} + \underbrace{h(x_j)[(1-v)(1-p)]}_{\text{prob. still winner \& no demand shift}} \times \underbrace{(1+\lambda)}_{\text{payoff in } t=2}$$

$$+\underbrace{h(1-x_j)[(1-v)p]}_{\text{prob. win } j' \text{ \& no demand shift}} \times \underbrace{(1+\lambda)}_{\text{payoff in } t=2}$$

Comparative static

x_j^* is increasing in v (= prob. other scientist wins)

- Invest more in active tech. when “competition” is stronger

- Examples of v in practice?
    - Actual competition from other scientists
    - Knowledge / skill / ability / etc.
    - Fixed costs

Social Planner's Expected Payoff
(it doesn't matter who wins)

$$\Pi(x_j) = h(x_j)[\underbrace{(1 + (1 - v)(1 - p)(1 + \lambda))}_{\text{private returns in } j} + \underbrace{v(1 - p)(1 + \lambda)^2}_{\text{social returns in } j}\ ]$$

$$h(1 - x_j)[\ \underbrace{[(1 - v)p(1 + \lambda))}_{\text{private returns in } j'} \ + \ \underbrace{vp(1 + \lambda)^2}_{\text{social returns in } j'}\ ]$$

Private vs. Social Optimum

Key result: $x_{j'}^{social*} > x_{j'}^{private*}$

- Social planner wants more effort in the alternative tech. (j') than scientist does

- Comparative static
  - Invest more in active tech. when "competition" is weaker, i.e., $\partial x_j^{social*} / \partial v < 0$
  - Recall, the opposite is true for the scientists' problem

$\Rightarrow$**Wedge between private and social optimum grows with "competition"!**

Counter-acting forces

that push against distortionary profit-seeking

- Adjustment costs
- Forecast (belief) differences
- Technology-specific competencies or preferences

- In other words, getting “stuck” in a certain field is great if your field happens to be valuable in the future!
- <u>Thesis I’d love to see</u>: *how close are observed adjustment costs of science to the socially optimal adjustment costs?*

Kortum's Comment
(in the same volume)

- In Acemoglu, early progress in non-active tech. quickly becomes superseded
- **Kortum: what about differing returns to scale?**

- How large are the dis-incentives from competition relative to the incentives from learning from more scientists?
    - *Is separating these forces from the aggregate returns to scale policy-relevant?*

The Economics of Science

an empiricist's view

Kyle R. Myers
kmyers@hbs.edu
Innovation Research Boot Camp, Summer 2024