

Discussion:

***“The Effect of Public Science on
Corporate R&D”***

by Arora, Belenzon, Cioaca, Sheer and Zhang

***“Measuring the Commercial
Potential of Science”***

by Masclans-Armengol, Hasan and Cohen

Ina Ganguli

University of Massachusetts Amherst and NBER

Universities as Engines of Growth

Innovation Policy Forum of the National Academies of Sciences, Engineering, and Medicine

May 15, 2024

Today's Question

*“Does federal funding of science,
especially through universities,
pay off in terms of commercial outcomes?”*

Today's Papers

Arora, Belenzon, Cioaca, Sheer and Zhang

- *How do scientific papers, patents, and PhD dissertations **produced at universities** impact corporate R&D?*
- *Does variation in **federal funding** across fields impact corporate R&D through these channels?*

Masclans-Armengol, Hasan and Cohen

- *What is the “commercial potential” of scientific papers **produced at universities**?*
- *Does university patenting of this knowledge impact its use by firms?*

#1: Arora, Belenzon, Cioaca, Sheer and Zhang

- Heroic data effort linking “public science” -- papers, patents and PhD theses -- to areas *relevant* to firms and their outcomes.
- To estimate the causal impact of these factors on firms, clever approach to create “randomness” in which firms to have access to “public science”
 - Federal agency R&D budgets are predicted by the share Republican/Democrat in congressional appropriations committees →
 - Leverage this variation in funding across across subfields to predict papers, patents, and theses →
 - This provides exogenous variation in firms’ access to them!

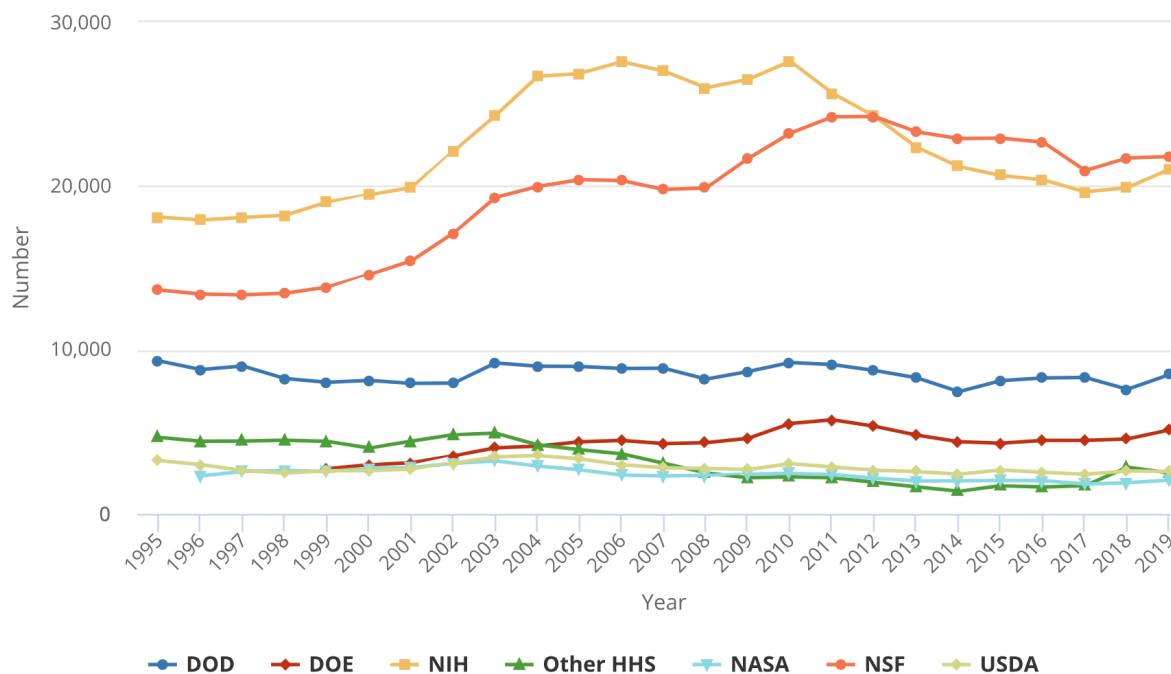
#1: Arora, Belenzon, Cioaca, Sheer and Zhang

- Key results:
 - An increase in PhD theses relevant to firms (similar to prior patents), increases firms' subsequent patents, publications, employment of scientists, and profits.
 - An increase in university patents decreases the firms' patents and publications.
 - An increase in university papers does not impact firms patents and publications

Comments

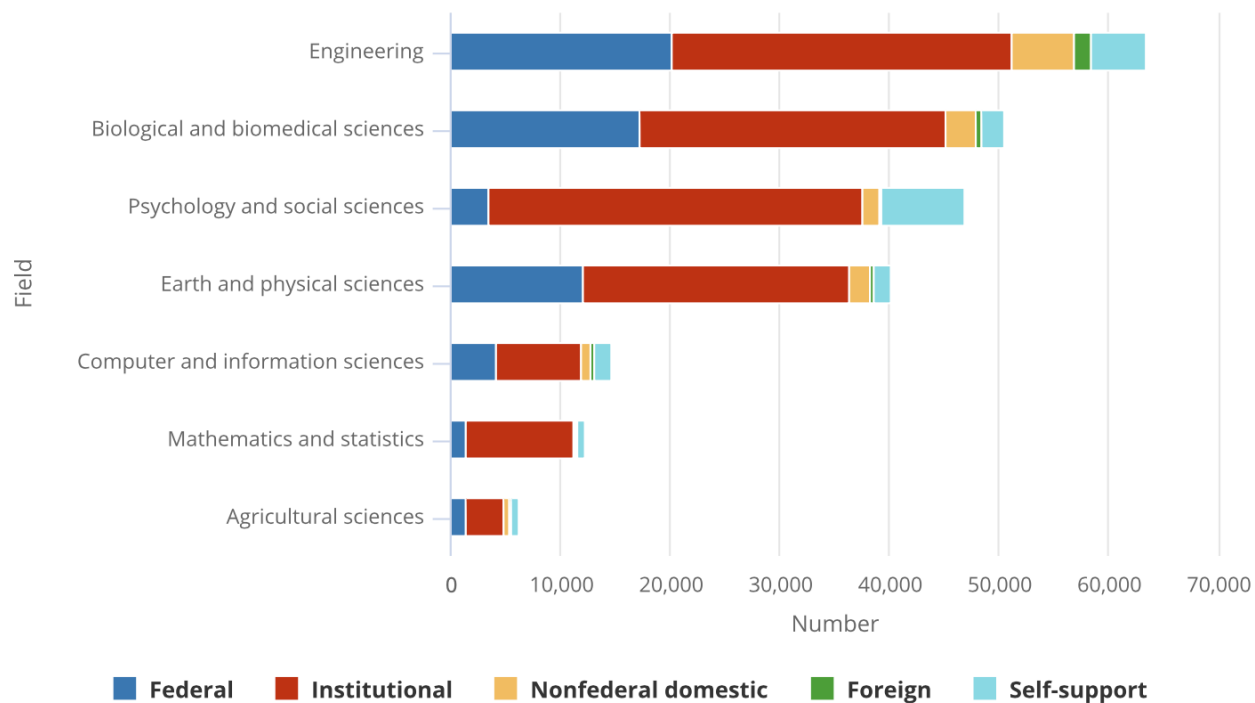
- First stage is interesting (and not trivial)!
 - Politics -- congressional subcommittee composition – influences funding and key inputs into university scientific production and firm outcomes.
- Are PhD theses human capital or public knowledge?
 - Theses are disclosed through ProQuest (data source) and often also published as scientific papers (before at the university, or after at the firm?)
 - Are the firms hiring the students after graduation (AMWS employment result)?
- How should we think about students NOT funded by federal grants?
 - Assume that “federal funding received by the advisor...affects the direction and content of the dissertation”

Graduate student federal funding, by agency



National Science Board, National Science Foundation. 2021. Science and Engineering Indicators 2022. Academic Research and Development. NSB-2021-3. Alexandria, VA. Available at <https://nces.nsf.gov/pubs/nsb20213/>

Graduate student funding, by source



National Science Board, National Science Foundation. 2021. *Science and Engineering Indicators 2022. Academic Research and Development. NSB-2021-3.* Alexandria, VA. Available at <https://nces.nsf.gov/pubs/nsb20213/>

#2: Masclans-Armengol, Hasan and Cohen

- Develop a measure of the “commercial potential” of scientific papers from universities using large language models and neural networks
- Commercial potential = “probability that a firm will view an article as contributing to the development of a marketable product or process” → “probability that a scientific article will be cited in a patent that is subsequently renewed”
- Importantly, they externally VALIDATE:
 - using rich data from one university TTO
 - analyzing commercial outcomes for 5.2 million academic papers and citations in renewed patents

#2: Masclans-Armengol, Hasan and Cohen

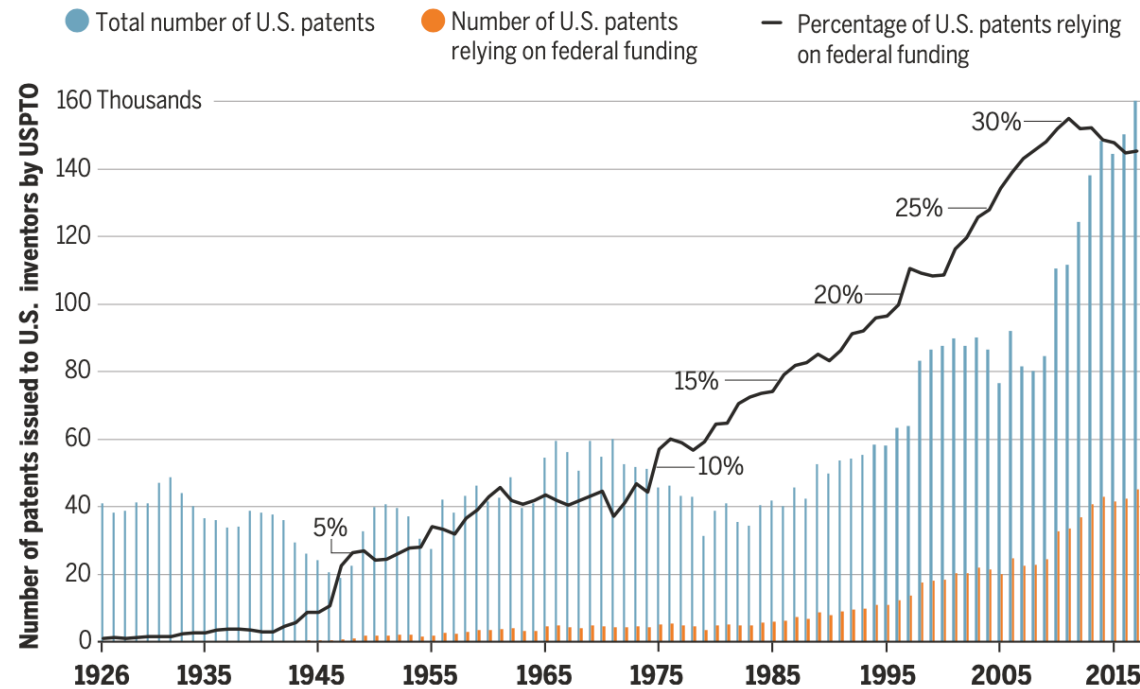
- Two applications:
 - Use a sample of scientific papers from one university and information on which are patented, show that 'high-commercial potential' papers that are patented are more likely to be cited than those not-patented.
 - Use a measure a university's commercialization reputation (commercial H-Index) to show that firms are less likely to cite 'high-commercial potential' papers produced by lower reputation universities.

Comments

- Great example of how LLM/NLP methods can be used to help us move forward on questions central to innovation
 - Important contribution to understanding the “realization gap” at universities – we usually cannot tell how much commercialize-able science is produced at universities that is NOT commercialized..
 - This measure can also be used to estimate how much federally-funded research *specifically* is not commercialized, but has potential (the question for today!)

Patentees increasingly depend upon federally supported research

Total granted U.S. patents by U.S. inventors (blue bars), and subtotal that rely on federal research (orange bars), and proportion of patents (black line = orange bars/blue bars) that rely on federally supported research.



Source: Fleming, L., Greene, H., Li, G., Marx, M. and Yao, D., 2019. Government-funded research increasingly fuels innovation. *Science*, 364(6446).

Comments

- How to think about using citations to papers in renewed patents as both a *predictor* of the commercial potential of a paper (in the LLM model) and as *an outcome of interest* for papers with commercial potential (in applications)?
- Can we assume that measure of commercial potential based on data available as of now will do well in the future (i.e. is it hard to predict the “long-tail”, breakthrough ideas)?

Thank you!

University of
Massachusetts
Amherst