Examining the Outcomes of Investments in Caltech Entrepreneurs (ETOILE)

PRINCIPAL INVESTIGATORS

Julia Lane (American Institutes for Research)

Paula Stephan (Georgia State University and NBER)

Jacques Mairesse (CREST-ENSAE, UNU-MERIT, and NBER)

OVERVIEW AND GOALS

This project is examining how the characteristics of principal investigators, research funding and the interaction of the two affect the dynamic structure of collaboration on scientific teams (their formation, evolution, and dissolution). What are the scientific and economic consequences of that collaboration? The project also seeks to understand the production of scientific output, modeling labs as firms. A goal is to understand the degree to which various inputs are substitutes or complements in production.

Using data from the California Institute of Technology derived from the STARMETRICS program (administrative data that draws on the financial records of institutions to document the flows of activities associated with science expenditures), the research team is:

1) Assessing the extent to which the formation and evolution of teams is driven by the presence of funding and how this formation differs by scientific research area
2) Examining the extent to which the formation and evolution of teams is driven by the presence of an entrepreneurial Principal Investigator and how this differs by scientific research area
3) Examining the extent to which various inputs are substitutes or complements in the production of scientific output.

STRUCTURE OF PROJECT

The major contribution of this research is to generate new measures of projects/teams as well as other inputs used in the production of scientific knowledge that go beyond measures that have been used previously.

1) Measure of team. The data can capture participants on funded projects, in different scientific areas over a long period of time.
2) Measure of PI characteristics. In addition to the standard STARMETRICS Level I data, we have richer information on the characteristics of Principal Investigators, from CV data.

3) Measures of other inputs used in the production process, such as equipment and certain types of materials
4) Scientific research areas. We use natural language processing to describe the research areas of interest.
5) Outputs. We will use the standard output sources (publications and patents) as well as look at the training of students and the speed of how outputs are generated. We also now expect to have information about student placement, firm startups and firm productivity from matches to Census data.

MAIN FINDINGS & SUCCESSES

1) Excellent insights from qualitative interviews of Caltech PIs
2) Uncovered rich information from expenditure data about purchases of materials, research inputs, services and big research equipment
3) We have begun to estimate multidimensional production functions at the project level
4) We have developed a prototype dashboard that reflects the linkages between scientific inputs and outputs

MAIN CHALLENGES

1) Disambiguation of vendor data and matching it to industry codes
2) Linking people to their outcomes – publicly available sources are either incomplete or messy
3) Setting up data structure that can both feed data to a dashboard and create an analytical database without losing the integrity of the data connections

DASHBOARD