Assessing the Impacts of NSF/TIP Investments on Regional Firms and Jobs

Context

The National Science Foundation's TIP Directorate is interested in assessing the impact of NSF/TIP investments on regional firms and jobs in specified technology areas. To this end, TIP is supporting a set of projects that are employing novel methods to provide such information. Assessing the impacts of research investments will be an on-going TIP activity, where novel methods and approaches will be explored for providing the most effective information.

Project

Industries of Ideas: A prototype system for measuring the effects of TIP investments on firms and jobs

PI: Jason Owen-Smith, University of Michigan; Co-PIs: Anna Harvey, Social Science Research Council/New York University, Bruce Weinberg, The Ohio State University.

Project Overview

This project is developing people-centric methods for following the movement of ideas from investments in research into the marketplace. This is done by identifying businesses that employ people with deep skills in emerging technology areas and developing early, never-before-available indicators that can provide alerts associated with potential impacts on the current and future workforce including, for example, the need for reskilling, upskilling, and new skill acquisition.

Investments in research, science, and technology are generating new "idea industries" that do not align neatly with traditional measures of industries and scientific fields. This project seeks to establish the foundation for a new, rapidly implementable, conceptual and empirical approach to tracing how these idea industries are formed by detecting flows of research-trained people into jobs and firms and the effects of those investments on the entire workforce affected by AI/EV investments.

Longitudinal data drawn from publication, grant, university and state administrative data will populate a prototype dashboard to describe research investments and their larger implications. Extensive stakeholder outreach will ensure broad accessibility and usefulness to a variety of stakeholders.

Approach

The pilot effort focuses on a single state, viz., the State of Ohio, and two initial technology areas, viz., Artificial Intelligence (AI) and Electric Vehicles (EV), to demonstrate the feasibility and value of the approach in performing rich descriptive analysis of the employment and workforce effects of science and technology investments.

Data types and Accessibility

The project uses data from multiple streams including public data as well as restricted-use data, creating a data mosaic for "ideas industries." A project dashboard will enable integrated data access to a broad range of users/stakeholders to help understand and explain the effects of TIP and NSF investments in emerging technologies on employers and jobs.

A 'tiered data access' scheme will enable access to different data that require different levels of authentication and permission. These include:

- Public release data. Public data such as the QCEW or the NSF HERD survey, which will be used for comparison and contextualization, will be readily and easily accessible to all potential research users.
- Privacy-disclosed results. Descriptive data from the pilot will be publicly available through the prototype dashboard and other mechanisms defined by stakeholder needs, after privacy disclosure review in accordance with all legal and ethical requirements of state and university data providers.
- Restricted data accessible through the project team. Access to research micro-data used in the project (viz., UMETRICS) will be possible under a Data Use Agreement (DUA) with the University of Michigan's Institute for Research on Innovation and Science (IRIS).
- Restricted data accessible through other entities. Access to workforce data, such as from the State of Ohio, is governed by the requirements of the state and relevant agencies which have established a formal approval process. Those seeking to access these data may do so through the Ohio Longitudinal Data Archive (ODLA).

Future Directions

- Future expansions of this effort include expanding the scale to include data from more states, universities, and technologies.
- Plans for expansions in scope will systematically include education data to characterize workforce skills and gaps (Figure 1, Step 5).
- The pilot project will result in descriptive information. Work with relevant research communities will yield key insights about needs for causal estimation in later project extensions.
- The scientific investment data in the pilot will make use of demographic data imputed from individual names and other data features using standard algorithmic techniques to describe the characteristics of the university research workforces supported by NSF, TIP, and other research investments. UMETRICS vendor data linked to business data from multiple sources will allow identification of women-, minority-, and veteran-owned businesses that are part of relevant research supply chains. State workforce data include no systematic information about individual demographic characteristics, so initial descriptions of workforce and employer information in terms of underrepresented groups will be limited by data sufficiency issues in this pilot. Planned expansions in scope to include educational data and potentially other restricted data (such as the NCSES Survey of Earned Doctorates, which is currently being linked to UMETRICS) will allow much more detailed and nuanced analysis on these dimensions in future project expansions.

Appendix A

Figure 1 describes the pilot pipeline including potential expansions to include educational data.

Measuring AI/EV investments and Identifying AI/EV researchers

- Step 1 identifies specific research investments and outcomes by applying machine learning models to identify research grants and scientific publications that pertain to each technology area.
- Step 2 identifies the people and firms funded by those research investments. This step uses detailed university administrative data included in the Universities Measuring the Effects of Research on Science, Technology, and Competitiveness (UMETRICS) data to find all related grants from federal, industry and non-profit sources that involve AI and EV researchers from Step 1. Step 2 can thus uncover all the people (faculty, staff, students, and trainees) that work on those projects as well as the non-academic vendors that supply goods and services to them. The scientific investment data that results will trace the cumulative impact of scientific investments in each field and permit analysis of the academic, scientific, and near-term economic spillovers associated with each.

Identifying AI/EV-intensive Firms

- Step 3 expands the people-centered approach to state workforce data using research-trained individuals and relevant research suppliers from Step 2 as a "tracing mechanism" to identify all employers who are active users and developers of AI/EV. This will be accomplished in two ways: first, State of Ohio workforce data will be used to identify employers that hire people who worked as students, staff, trainees, or faculty on relevant grants at universities in the state. Second, data from UMETRICS will be used to identify all the vendors supplying high-technology AI/EV research at those institutions. Hiring highly trained (often Ph.D. level) people with demonstrable research experience with key technologies provides an important indicator of an employer's engagement with the technology itself. By focusing initially on people hired with deep skills in particular technological areas, this approach provides an industry/application-agnostic indicator of the employers actively and directly engaged in relevant R&D and those who are developing or applying technology-based tools and methods. As a result, the project will describe the wide range of businesses engaged with these technologies in a fashion that directly connects to NSF and other research investments and that facilitates key comparisons across funders, fields, and established industries.

- Step 4 expands the analysis of state workforce data to describe all of the people who have worked at organizations that are developing or using emerging technologies, thus offering a snapshot of the complete workforce of employers engaged in AI/EV work. These data will also enable numerous comparisons between the workforce of idea industries defined by this approach, and that identified using more traditional classification-based approaches such as NAICS codes. The resulting longitudinal employer and workforce data will be combined with scientific investment data to produce a robust and evolving picture of investments in emerging technologies and their larger impact on employers, jobs, industries, and states.

Figure 1. The Flow from AI/EV Research Investments to Jobs