

IRIS Research Bibliography

PUBLISHED PAPERS

Science Funding and Short-Term Economic Activity

Weinberg BA, Owen-Smith J, Rosen RF, Schwarz L, McFadden Allen B, Weiss RE, & Lane JI

***Science* 4 April 2014**

Vol. 344(6179), pp. 41-43 DOI: 10.1126/science.1250055

<http://www.sciencemag.org/content/344/6179/41.full>

Abstract

There is considerable interest among policy-makers in documenting short-term effects of science funding. A multiyear scientific journey that leads to long-term fruits of research, such as a moon landing, is more tangible if there is visible nearer-term activity, such as the presence of astronauts. Yet systematic data on such activities have not heretofore existed. The only source of information for describing the production of most science is surveys that have been called “a rough estimate, frequently based on unexamined assumptions that originated years earlier.

New Linked Data on Research Investments: Scientific Workforce, Productivity, and Public Value

Lane JI, Owen-Smith J, Rosen RF, & Weinberg BA

***Research Policy* December 2014**

Vol. 44(9), pp. 1659-1671 DOI: 10.1016/j.respol.2014.12.013

<http://www.sciencedirect.com/science/article/pii/S0048733315000025>

Abstract

Longitudinal micro-data derived from transaction level information about wage and vendor payments made by Federal grants on multiple US campuses are being developed in a partnership involving researchers, university administrators, representatives of Federal agencies, and others. This paper describes the UMETRICS data initiative that has been implemented under the auspices of the Committee on Institutional Cooperation. The resulting data set reflects an emerging conceptual framework for analyzing the process, products, and impact of research. It grows from and engages the work of a diverse and vibrant community. This paper situates the UMETRICS effort in the context of research evaluation and ongoing data

infrastructure efforts in order to highlight its novel and valuable features. Refocusing data construction in this field around individuals, networks, and teams offers dramatic possibilities for data linkage, the evaluation of research investments, and the development of rigorous conceptual and empirical models. Two preliminary analyses of the scientific workforce and network approaches to characterizing scientific teams ground a discussion of future directions and a call for increased community engagement.

Wrapping It Up in a Person: Examining Employment and Earnings Outcomes for Ph.D. Recipients

Zolas N, Goldschlag N, Jarmin RS, Stephan P, Owen-Smith J, Rosen RF, McFadden Allen B, Weinberg BA, & Lane JI

***Science* 11 December 2015**

Vol. 350(6266), pp. 1367-1371

DOI: 10.1126/science.aac5949

<http://www.sciencemag.org/content/350/6266/1367.full>

Supplementary material: http://econ.ohio-state.edu/weinberg/Science-aac5949_Zolas-SM-PUBLISHED.pdf

Abstract

In evaluating research investments, it is important to establish whether the expertise gained by researchers in conducting their projects propagates into the broader economy. For eight universities, it was possible to combine data from the UMETRICS project, which provided administrative records on graduate students supported by funded research, with data from the U.S. Census Bureau. The analysis covers 2010–2012 earnings and placement outcomes of people receiving doctorates in 2009–2011. Almost 40% of supported doctorate recipients, both federally and nonfederally funded, entered industry and, when they did, they disproportionately got jobs at large and high-wage establishments in high-tech and professional service industries. Although Ph.D. recipients spread nationally, there was also geographic clustering in employment near the universities that trained and employed the researchers. We also show large differences across fields in placement outcomes.

STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data Linked to the 2010 Census

Buffington C, Cerf B, Jones C, & Weinberg BA

***American Economic Review* May 2016**

106(5), pp. 333–338 DOI: 10.1257/aer.p20161124

<https://www.aeaweb.org/articles?id=10.1257/aer.p20161124>

Abstract

Women are underrepresented in science and engineering, with the underrepresentation increasing in career stage. We analyze gender differences at critical junctures in the STEM pathway—graduate training and the early career—using UMETRICS administrative data matched to the 2010 Census and W-2s. We find strong gender separation in teams, although the effects of this are ambiguous. While no clear disadvantages exist in training environments, women earn 10% less than men once we include a wide range of controls, most notably field of study. This gap disappears once we control for women’s marital status and presence of children.

Why the U.S. Science and Engineering Workforce is Aging Rapidly

Blau D, & Weinberg BA

***Proceedings of the National Academy of Sciences*, 14 February 2017**

Vol. 114(15), 3879–3884 DOI: 10.1073/pnas.1611748114/-/DCSupplemental

<http://www.pnas.org/content/114/15/3879.short>

Abstract

The science and engineering workforce has aged rapidly in recent years, both in absolute terms and relative to the workforce as a whole. This is a potential concern if the larger number of older scientists crowds out younger scientists, making it difficult for them to establish independent careers. In addition, scientists are believed to be most creative earlier in their careers, so the aging of the workforce may slow the pace of scientific progress. The authors developed and simulated a demographic model, which shows that a substantial majority of recent aging is a result of the aging of the large baby boom cohort of scientists. However, changes in behavior have also played a significant role, in particular a decline in the retirement rate of older scientists, induced in part by the elimination of mandatory retirement in universities in 1994. Furthermore, the age distribution of the scientific workforce is still adjusting. Current retirement rates and other determinants of employment in science imply a steady-state mean age 2.3 years higher than the 2008 level of 48.6.

COMMENTARIES

- [Fix Incentives](#) (Nature: Perspective September 1, 2016)
Julia Lane
- [Who Feels the Pain of Science Research Budget Cuts?](#) (The Conversation/Salon March 29, 2017) Bruce Weinberg
- [A call to action to build social science data infrastructure](#) (Nature Human Behaviour April 7, 2017) Julia Lane
- [The social sciences need to build new foundations](#) (Significance Magazine June 9, 2017)
Julia Lane
- [Watching the players, not the scoreboard](#) (Nature: Comment November 2, 2017)
Julia Lane
- [Tax bill would imperil nation's innovation, future](#) (Columbus Dispatch December 14, 2017)
Bruce Weinberg, Jason Owen-Smith, and Julia Lane

FORTHCOMING

Research Funding and Regional Economies

Goldschlag N, Bianchini S, Lane JI, SanMartin Sola J, & Weinberg BA

National Bureau of Economic Research Working Paper No, 23018, January 2017

DOI: 10.3386/w23018 <http://www.nber.org/papers/w23018>

Proceedings of the National Academy of Sciences submission

Abstract

The practice of publicly funded science is global; research collaborations span countries; researchers move across continents. Yet public support of research typically relies on the notion that universities are engines of economic development, and that university research is a primary driver of R&D intensive high wage localized economic activity. This paper establishes that it is possible to use transactional data to capture the dynamics between research grants and associated purchases from regional suppliers in both Europe and the United States. Here we use new micro-data from three countries – France, Spain and the United States – to examine one mechanism whereby such economic activity is generated, namely purchases from

regional businesses. We show that grant funds are more likely to be expended at businesses physically closer to universities than at those farther away. In addition, if a vendor has been a supplier to a grant once, that vendor is subsequently more likely to be a vendor on the same or related grants. Firms behave in a way that is consistent with the notion that propinquity is good for business; if a firm supplies a research grant at a university in a given year it is more likely to open an establishment near that university in subsequent years than other firms.

Linking in a Big Data World

Chang W, Emad A, Lane JI, Togle J, & Weinberg BA
Science Policy Forum submission

Abstract

The increased availability of new types of data means that both social science researchers and statistical agencies are interested in finding new ways of linking survey and other datasets together. A major challenge, however, is that unique identifiers are typically not present, nor are gold standard datasets. This paper describes the process of linking a well curated survey, the Survey of Earned Doctorates to a new source of transaction data which featured all of these challenges. We were able to show that machine learning approaches can successfully be used to improve link quality, and that great use can be made of data that are available in a subset of files. We also extended the literature that uses output measures - analytical utility – in addition to input measures – match quality – as metrics of the success of the linkage effort.

Federal Funding of Doctoral Recipients: Results from New Linked Survey and Transaction Data

Chang W, Cheng W, Jones C, Lane JI, & Weinberg BA
Journal of Official Statistics submission

Abstract

This paper has combined a new administrative dataset (UMETRICS) with a major survey dataset (SED) to provide new insights into the links between federal funding structure and the composition of the doctorally trained workforce. The contribution is substantial, since current surveys do not provide information about the extent or intensity of federal funding, and no individual agency, looking at its own data, can document how many doctoral recipients are

supported by federal funds. We find large differences in funding patterns across federal funding agencies in terms of the number and disciplinary training of doctoral students. Federal funding is also strongly related to career pathways. Much more can be done; the analysis draws on data for only two major research universities for which long historical data exist. As data from the new Institute for Research on Innovation and Science expands across more institutions and longer time periods, the analysis can become more generalized. It can also be extended to examine the links to non-federal sources of funds, as UMETRICS data is expanding to include that information for partner institutions. And with those links, more causal analysis can be added to the descriptive characterization provided in this paper.

Research Experience as Human Capital in New Business Outcomes

Lane JJ, Jarmin RS, Goldschlag N, & Zolas N

American Economic Association Meeting, Philadelphia, January 2018

Session: New Measures of Human Capital and Their Application (moderated by Bruce Weinberg)

Forthcoming in CRIW volume on *The Measurement and Diffusion of Innovation*

Abstract

Human capital is typically cited as an important contributor to the survival, growth and innovative activity of new businesses. This paper contributes to the literature by both developing novel measures of human capital and examining the link between those measures and the outcomes of young firms. It builds on several strands of the literature which emphasize the importance of employee workplace experience as a dimension of human capital. It shows that the effects of work experience differ substantially by where an employee worked and is valued differently by firms in different sectors. This is particularly true for research experience, which is consistent with the notion that on the job training in complex tasks should be valuable to firms with complex production technologies. This paper will be included as a book chapter in the NBER CRIW volume on *The Measurement and Diffusion of Innovation* (Carol Corrado, Javier Miranda, Jonathan Haskel, and Daniel Sichel, eds., University of Chicago Press, forthcoming)

Research Universities and the Public Good: Discovery for an Uncertain Future

Owen-Smith, Jason

Stanford University Press. Forthcoming, September 2018

Abstract

This book uses my original research, current and historical case studies of important discoveries, and analysis of secondary sources to argue that a small number of research intensive universities (about 3% of all post-secondary institutions) serve a unique and essential role for the U.S. and the world. They are a form of “social insurance” that keeps our nation poised to respond to unexpected challenges and to develop new opportunities. The way that they do this is too little understood and this book seeks to make it clear for a broad readership. This is important because current policies, debates, and trends stand to damage or destroy our research universities’ ability to accomplish this important work.

Proximity and Economic Activity: An Analysis of Vendor-University Transactions

Nathan Goldschlag, Julia Lane, Bruce Weinberg and Nikolas Zolas
Forthcoming in *Journal of Regional Science*

Abstract

This paper using transaction based data to provide new insights into the link between the geographic proximity of businesses and associated economic activity. It contributes to the literature by developing both two new measures of distance and a set of stylized facts on the distances between observed transactions between vendors and customers for a research intensive sector – universities. We show that spending on research inputs is more likely to be expended at businesses physically closer to universities than those farther away. That relationship is stronger for High Tech and R&D performing businesses than businesses in general, which is consistent with theories emphasizing the role of tacit knowledge. We find that firms behave in a way that is consistent with the notion that propinquity is good for business: if a firm supplies a project at a university in a given year, it is more likely to open an establishment near that university in subsequent years than other firms. We also investigate the link between transactional distance and economic activity and show that when a vendor has been a supplier to a project at least one time, that vendor is subsequently more likely to be a vendor on the same or related project.

WORKING PAPERS

Evaluating Author Name Disambiguation for Digital Libraries: A Triangulation Approach

Kim, Jinseok

Submitted for publication (2018).

Abstract

Equipped with advanced computing techniques, scholars have disambiguated author names in whole digital libraries and tested their performances in various ways. The purpose of this study is to propose a triangulation approach that author name disambiguation for digital libraries can be better evaluated when its performance is assessed on multiple labeled datasets with comparison to baselines for diverse ambiguity dimensions. To illustrate the proposed approach, accuracy of author name disambiguation in DBLP's 3.7M records is evaluated on three types of labeled data containing 5,000 to 6M disambiguated names. Results show that the triangulation method can provide a more holistic, granulated understanding of a disambiguation method's performance than common evaluation practices in prior studies. With the review of strengths and weaknesses of the proposed approach, this study calls for further discussion about consistent frameworks and methodologies for evaluating author name disambiguation so that findings from a variety of studies can be synthesized to produce insights for improving name ambiguity resolution for fast-growing digital libraries.

Kim, J. & Kim, J. (2018). The Impact of Imbalanced Training Data on Machine Learning for Author Name Disambiguation. Submitted for publication.

In supervised machine learning for author name disambiguation, negative training data are often dominantly larger than positive training data. This paper examines how the ratios of negative to positive training data can affect the performance of machine learning algorithms to disambiguate author names in bibliographic records. On multiple labeled datasets, three classifiers – Logistic Regression, Naïve Bayes, and Random Forest – are trained through representative features such as author name, coauthor names, and title words extracted from the same training data but with various positive-to-negative training data ratios. Results show that increasing negative training data can improve disambiguation performance but with a few percent of performance gains and sometimes degrade it. Logistic Regression and Naïve Bayes learn optimal disambiguation models even with a base ratio (1:1) of positive and negative training data. Also, the performance improvement by Random Forest tends to quickly saturate roughly after 1:10 ~ 1:20. These findings imply that contrary to the common practice using all training data, name disambiguation algorithms can be trained using part of negative training data without degrading much disambiguation performance while increasing computational efficiency. This study calls for more attention from author name disambiguation scholars to methods for machine learning from imbalanced data.

Automatic Generation of Labeled Data for Author Name Disambiguation: An Iterative Blocking Method.

Kim, Jinseok & Owen-Smith, Jason
Submitted for publication (2018).

Abstract

Many author name disambiguation studies have relied on hand-labeled truth data that are very laborious to generate. This paper shows that labeled data can be automatically generated using information features such as email address, coauthor names, and cited references that are available from publication records. For this purpose, high-precision rules for matching name instances on each feature are learned using an external-authority database. Then, selected name instances in target ambiguous data go through the process of pairwise matching based on the learned rules. Next, they are merged into blocks by a generic entity resolution algorithm. The blocking procedure is repeated over other features until further merging is impossible. Tested on an example of 26,566 name instances, this iterative blocking produced accurately labeled data with near perfect accuracy (pairwise F1 = 0.99). In addition, the labeled data represented the population data of 227K name instances in terms of name ethnicity and co-disambiguating name group size distributions. Several challenges are discussed for applying this method to resolving author name ambiguity in large-scale scholarly data.

How Universities Organize Science

Fisher, Jacob C. & Owen-Smith, Jason
Submitted for publication (2018)

Abstract

Although the results of science -- publications and patents -- have received considerable attention, little work to date has considered how the production of science is organized. Using a unique dataset on grant payments to faculty, staff, and trainees within 23 universities, we explore how universities approach a similar task, developing research findings, in different ways. Drawing on organizational theory that suggests that work is accomplished through a network of collaborations, we examine two complementary processes that cause the organization of science to differ between universities. First, administrators and grantors can control the number and occupation of people involved in the network. Second, individual faculty members can control the specific collaboration relationships, both between faculty members, and among staff and trainees who receive funding from particular grants. In network terms, administrators and grantors control the vertices, and individual faculty control the edges between them. We find that the influence of administrators and grantors is most visible along

two dimensions: the amount of funding awarded by NIH, and the ratio of trainees to staff. A cluster analysis demonstrates that individual faculty staff grants in one of six ways, which depend on the scale of the grant and the faculty member's preferences. We find that between-university differences in the connectivity of the network can largely be explained by differences in scale, differences in clustering can be explained by faculty preferences, but overall differences in structure of the networks cannot be well-explained by either scale or collaboration preferences.

The link between R&D, human capital and business startups

Goldschlag N, Jarmin RS, Lane JI, & Zolas N

American Economic Association Meeting, Chicago, January 2017

Session: Using Data Science to Examine the Link Between University R&D and Innovation
(moderated by Julia Lane)

Abstract

The reason for the secular decline in entrepreneurship is not well understood. It is evident in all sectors of the economy and almost all regions. One approach to stimulating innovation and entrepreneurship has been to increase investments in science: the U.S. federal government contributed nearly \$38 billion for university-based research in Science, Technology, Engineering, and Mathematics (STEM) in 2014. However, there has historically been little evidence about the links between investments in university research and innovation - largely because surveys cannot capture the complex ways in which scientific ideas are created, transmitted and adopted.

This paper examines the relationship between the funding of research teams - in terms of structure, field and type of funding - and the subsequent propensity of members of those teams to start up businesses. It also examines the subsequent survival and productivity growth of those startups.

The work is now possible because of a new data infrastructure resulting from collaborations between the Census Bureau's Innovation Measurement Initiative, the National Science Foundation and the Institute for Research on Innovation and Science at the University of Michigan. The infrastructure links universe data on all people employed on research grants, their funding, and their economic and scientific activities.

This paper is the first to directly trace the pathways from the bench to the workplace at a large scale, using universe data from 25 universities covering about 25% of federal university based R&D. It is the first to use universe data on workers (the LEHD data) to draw comparison groups

of individuals employed both within the university and from other R&D intensive businesses. And it is the first to use universe data on business startups to compare the dynamics of university sourced entrepreneurship with other types of entrepreneurship.

Pathways to Production

Barth E, Davis J, Marschke G, Wang A, Zhou S

American Economic Association Meeting, Chicago, January 2017

Session: Using Data Science to Examine the Link Between University R&D and Innovation
(moderated by Julia Lane)

Abstract

Science funding agencies often require researchers to demonstrate their project's prospects for "development of a diverse, globally competitive STEM workforce," "increase[d] partnerships between academia, industry and others" (NSF, 2016), and other goals beyond the creation of scientific knowledge. This paper attempts to measure these wider impacts of scientific research.

We use the newly created Census data infrastructure that links university grant transaction data to Longitudinal Employer-Household Dynamics (LEHD) data to map employment linkages between universities and industry. First, we ask, what are the flow rates of new STEM workers—post-docs and recent doctorates—into research intensive firms, industries, and regions? startups and established firms? high- and low- productivity firms? local and out-of-state employment?

Second, we estimate the impacts of and returns to university-based human capital accumulation by STEM workers. The sudden increase in science funding under the American Recovery and Reinvestment Act of 2009 (ARRA) increased demand in the academic sector for post-graduate researchers, both lengthening existing post-graduate research engagements in universities, and increasing the likelihood that recent graduates, especially doctorates, obtain post-graduate employment in universities. We estimate the impact of increased university-based research training on career paths, including the likelihood of obtaining a faculty post, and for researchers who enter industry, which firms they match to, and their wage outcomes.
Third, we investigate the extent to which a firm placement depends on the history of previous placements from the same university. Such a correlation could be evidence of "hiring chains", or of specific knowledge links between the research and teaching at a specific university and the production technology of particular firms. The hiring patterns we uncover between universities and industry reveal important features of the labor market for specialized skills, and

increase our understanding of how university research contributes to the diffusion of new ideas in the economy.

Financial Advice and the Entrepreneurial Spillovers of Basic Research

Dacunto F, & Yang L

American Economic Association Meeting, Chicago, January 2017

Session: Using Data Science to Examine the Link Between University R&D and Innovation
(moderated by Julia Lane)

Abstract

We test for the effect of informal financial advice on the establishment and subsequent performance of entrepreneurial ventures that commercialize the results of basic research. To this aim, we construct a unique data set that includes: (i) the characteristics of the faculty recipients of federally-funded grants across 10 large U.S. universities, which produce innovation that can be commercialized through the establishment of startups; (ii) the likelihood that recipients establish a non-employer venture (iLBD) or an employer venture (LEHD), as well as the job growth characteristics of these ventures; and (iii) the network of neighbors in the locations where the recipients' reside, including the occupation titles and demographics of the neighbors (ACS/Decennial Census). We use the presence of financial-sector employees among the faculty's network members (spouses or neighbors) to test for the effect. We compare faculty grant recipients in similar areas of research, obtaining grants of similar sizes in the same rounds of funding, and at similar stages of their academic careers, but belonging to networks with different levels of exposure to informal financial advice from family and friends. Financial advice from one's social network is informal because advisers are not paid fees for providing their service. Therefore, the paper broadly tests for whether advice is a positive externality of one's social networks, which is valuable to the individual entrepreneurs as well as to economic growth.

Research Funding and Subsequent Entrepreneurship: The Role of Underrepresentation

Buffington C, Harris B, Feng F, & Weinberg BA

American Economic Association Meeting, Chicago, January 2017

Session: Using Data Science to Examine the Link Between University R&D and Innovation
(moderated by Julia Lane)

Abstract

Federal funding affects both who does research, and the environment in which research is done. In a recent study, 6 in 10 female doctoral recipients had been supported by federal research funds, compared to 7 in 10 male doctoral recipients. Federal funding also appears to be highly correlated with the pipeline of researchers going into different fields; particularly into R&D fields and the decision to pursue postdoctoral fellowships.

This paper uses rich new Census Bureau data linked to detailed information on the individuals supported by research funding to examine the effect of both the type and structure of federal funding on the outcomes of underrepresented students. It makes use of rich measures on student characteristics, including their race, gender, place of birth, marital status and presence of children. It constructs new network theoretic measures of team environment, based on the characteristics of all individuals working together on research grants. It also includes information about household and family structure in the model. It also examines two types of outcome measures - placement in R&D performing, high technology or young and small firms - as well as the propensity of underrepresented groups to start up businesses.

BOOKS & BOOK CHAPTERS

Big Data and Social Science: A Practical Guide to Methods and Tools

Foster I, Ghani R, Jarmin RS, Kreuter F, & Lane JI (eds.)

Chapman and Hall/CRC, August 9, 2016

ISBN 9781498751407

Abstract

Big Data and Social Science: A Practical Guide to Methods and Tools shows how to apply data science to real-world problems in both research and the practice. The book provides practical guidance on combining methods and tools from computer science, statistics, and social science. This concrete approach is illustrated throughout using an important national problem, the quantitative study of innovation. The text draws on the expertise of prominent leaders in statistics, the social sciences, data science, and computer science to teach students how to use modern social science research principles as well as the best analytical and computational tools. It uses a real-world challenge to introduce how these tools are used to identify and capture

appropriate data, apply data science models and tools to that data, and recognize and respond to data errors and limitations.

Chapter 8 Networks: The Basics

Owen-Smith J, in *Big Data and Social Science* pp. 215-240

Foster I, Ghani R, Jarmin RS, Kreuter F, & Lane JI (eds.)

Chapman and Hall/CRC, August 9, 2016

ISBN 9781498751407

Abstract

Noted sociologist and network theorist Jason Owen-Smith provides a primer on network theory, including details on network measures and components.

Measuring the Economic Value of Research: The Case of Food Safety

Husbands Fealing K, Lane JI, King J, & Johnson SR (eds.)

Cambridge University Press, 2017

ISBN 9781316671788

Abstract

This innovative study demonstrates new methods and tools to trace the impact of federal research funding on the structure of research, and the subsequent economic activities of funded researchers. The case study is food safety research, which is critical to avoiding outbreaks of disease. The authors make use of an extraordinary new data infrastructure and apply new techniques in text analysis. Focusing on the impact of U.S. federal food safety research, this book develops vital data-intensive methodologies that have a real world application to many other scientific fields.

PRESS

- [*Biologists lose out in post-PhD earnings analysis*](#) (Nature: News December 10, 2015)
- [*Where new PhD grads find work — and who earns the most with their degree*](#) (Washington Post December 10, 2015)

- [*PhDs pay: study reveals economic benefit of funding doctorates*](#) (Times Higher Education December 10, 2015)
- [*Science and math PhDs earn about \\$65,000 — more than double what arts majors do*](#) (Vox December 11, 2015)
- [*ProQuest Dissertation Database Provides Critical Information for Research Projects Across the US*](#) (PR Newswire March 22, 2016)
- [*Assessment: Academic return*](#) (Nature May 4, 2016)
- [*There's a huge gender pay gap for STEM careers — just one year after graduation*](#) (Vox May 11, 2016)
- [*Facing Skepticism, colleges set out to prove their value*](#) (PBS Newshour January 22, 2016)
- [*The Price of Doing a Postdoc*](#) (Science: Share January 10, 2017)
- [*Trump Administration Proposes Big Cuts in Medical Research*](#) (NPR Health Shots March 16, 2017)
- [*Communicating the Value of University Research When Science is Under Attack*](#) (Inside Higher Ed April 6, 2017)
- [*The Looming Decline of the Public Research University*](#) (Washington Monthly September/October 2017)