

Data Science and Social Science

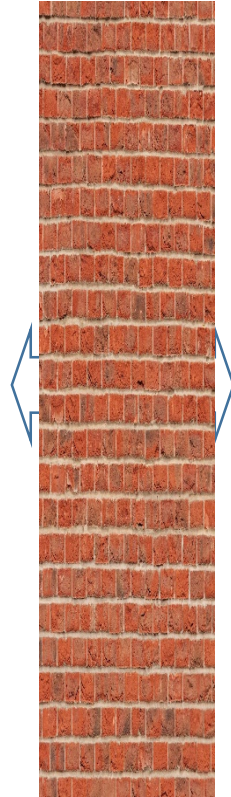
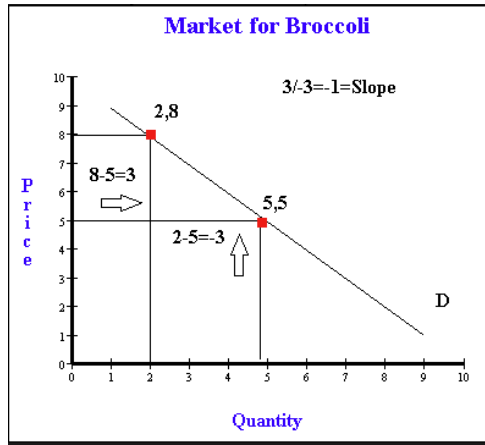
“More than the sum of its parts”

Kosali Simon

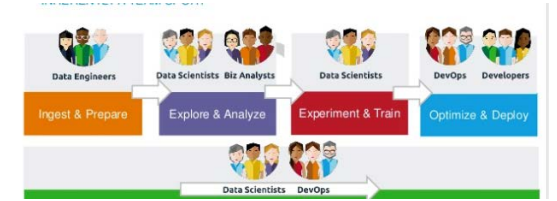
SPEA-IUB/NBER

Other IU affiliations: Kelley/Data Science/Regenstrief

Economics



Data Science



MODERN DATA SCIENTIST

Data Scientist, the saviest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Bayesian modeling
- Bayesian inference
- Generalized linear models
- Linear regression
- Logistic regression
- Dimensionality reduction
- Dimensionality reduction
- Dimensionality reduction
- Dimensionality reduction

PROGRAMMING & DATABASE

- Computer science fundamentals
- Database management systems (DBMS)
- Database design and modeling
- Database design and modeling
- Database design and modeling
- Database design and modeling
- Database design and modeling
- Database design and modeling
- Database design and modeling
- Database design and modeling

BUSINESS & COMMUNICATION

- Business analysis
- Business analysis
- Business analysis
- Business analysis
- Business analysis
- Business analysis
- Business analysis
- Business analysis
- Business analysis
- Business analysis



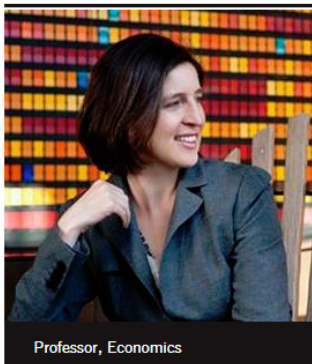
Not entirely separated, but different in core principles

Examples where SS and DS
interact

Susan Athey
Michael Luca

Working Paper 25064
<http://www.nber.org/papers/w25064>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2018



Susan Athey

STANFORD
BUSINESS
SCHOOL

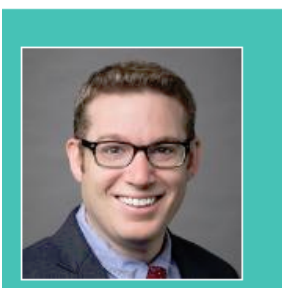
The Economics of Technology Professor

Professor of Economics (by courtesy), School of Humanities and Sciences
Senior Fellow, Stanford Institute for Economic Policy Research

Academic Area: Economics

Additional Administrative Titles

Co-Director, Digital Business: Data, Decisions & Platform Strategy Initiative



Michael Luca

HARVARD
BUSINESS
SCHOOL

Lee J. Styslinger III Associate Professor of Business Administration

Michael Luca is Lee J. Styslinger III Associate Professor of Business Administration at Harvard Business School. Professor Luca works closely with companies and cities to help them become more data-driven, and has ongoing collaborations with Yelp, Facebook, the UK government, and the City of Boston, in addition to other partners.

 [Print Entire Profile](#)

 [MORE](#)

ABSTRACT

As technology platforms have created new markets and new ways of acquiring information, economists have come to play an increasingly central role in tech companies – tackling problems such as platform design, strategy, pricing, and policy. Over the past five years, hundreds of PhD economists have accepted positions in the technology sector. In this paper, we explore the skills that PhD economists apply in tech companies, the companies that hire them, the types of problems that economists are currently working on, and the areas of academic research that have emerged in relation to these problems.

The job market for economists in tech sector

- Amazon: 150 PhD Economists hired in last 5 yrs
 - Compare to Bell Labs 1968-83 economist team (30), journal: (RAND)
 - Small legal expert witness industry (mainly as consultants)
 - Other than that, this is unprecedented non-academic interactions outside of policy evaluation
- Recruited through regular phd market
 - 2/3rds as many tech cos as policy schools currently hiring economists

Figure 2: The number of positions in tech companies, policy schools, business schools, and economics departments

	<i>Tech companies</i>	<i>Policy schools</i>	<i>Business schools</i>	<i>Economics departments</i>
<i>Feb 2017-Jan 2018</i>	21	34	162	194
<i>Feb 2016-Jan 2017</i>	20	23	149	199
<i>Feb 2015-Jan 2016</i>	18	31	150	218
<i>Feb 2014-Jan 2015</i>	15	26	133	232

The job market for economists in tech sector

- Working directly on business problems, although some in research centers
- Business Schools:
 - Hiring economists publishing in tech sector
 - MBA courses adding DS and Econ methods

Impact on Each Field

- Causal perspective added to DS/Tech field
 - Economics has developed strong toolkit for causal analysis
 - Theory helps anticipate intended and unintended consequences, feedback loops
 - Eg economists showed race bias interest in some Airbnb user functions
- New/better data & data tools added to Economics field
 - Economics has been slow to warm to ML/AI
- New questions added to Economics
 - Economics of digitization
 - E.g. how does AI and “bigdata” on consumers affect market structure and regulation?
 - How does social/organizational decision making change?

Impact on Each Field

- Causal perspective added to DS/Tech field
 - Economics has developed strong toolkit for causal analysis
 - Theory helps anticipate intended and unintended consequences, feedback loops
 - Eg economists showed race bias interest in some Airbnb user functions
- **New/better data & data tools added to Economics field**
 - Economics has been slow to warm to ML/AI
- New questions added to Economics
 - Economics of digitization
 - E.g. how does AI and “bigdata” on consumers affect market structure and regulation?
 - How does social/organizational decision making change?

Impact on Each Field

- Causal perspective added to DS/Tech field
 - Economics has developed strong toolkit for causal analysis
 - Theory helps anticipate intended and unintended consequences, feedback loops
 - Eg economists showed race bias interest in some Airbnb user functions
- New/better data & data tools added to Economics field
 - Economics has been slow to warm to ML/AI
- **New questions added to Economics**
 - **Economics of digitization**
 - E.g. how does AI and “bigdata” on consumers affect market structure and regulation?
 - How does social/organizational decision making change?

Data Science Institute

ABOUT

CENTERS

ACADEMICS

RESEARCH

CENTERS

Computing Systems for Data-Driven Science

Cybersecurity

Data, Media & Society

Financial & Business Analytics

Computational Social Science

The [Working Group on Computational Social Science](#) at Columbia University, in partnership with the Institute for Social and Economic Research and Policy (ISERP) and the Data Science Institute (DSI) explores profound and urgent questions about how the information revolution will transform society as a whole.

BERKELEY

Institute for
Data Science



ABOUT ▾

PROJECTS

PEOPLE

P

Social Science

Environment and Society: Data Sciences for the 21st Century (DS421)

Two versions of DS & Economics Interactions

- Wrangling/cleaning/analyzing/visualizing data
- Vs
- Technical tools, data architectures, processes from CS: ML random forests, optimizing algorithms in new ways, Apache Spark data platforms, NLP , AI to automate literature searches (e.g. Meta)



AMERICAN
ECONOMIC
ASSOCIATION



American Economic Journal: Economic Policy

American Economic Journal: Economic Policy publishes papers covering a range of topics, the common theme being the role of economic policy in economic outcomes.

American Economic Journal: Economic Policy 2018, 10(3): 154–192
<https://doi.org/10.1257/pol.20160248>

Strategic Formulary Design in Medicare Part D Plans[†]

By KURT LAVETTI AND KOSALI SIMON*

The design of Medicare Part D causes most beneficiaries to receive fragmented health insurance, with drug and medical coverage separated. Fragmentation is potentially inefficient since separate insurers optimize over only one component of healthcare spending, despite complementarities and substitutabilities between healthcare types. Fragmentation of only some plans can also lead to market distortions due to differential adverse selection, as integrated plans may use drug formularies to induce enrollment by patients that are profitable in the medical insurance market. We study the design of insurance plans in Medicare Part D and find that formularies reflect these two differences in incentives. (JEL D82, G22, H51, I13, I18, L65)

**Research to detect
company behavior in
response
to regulations
(using universe data)**

More sophisticated examples



Synthetic Control Using Lasso

Choose weights to satisfy: $\operatorname{argmin}_{\beta} \left\{ \frac{1}{2N} \sum_{t=1}^{T_0} (Y_{1t} - \sum_{s=2}^S \beta_s Y_{st})^2 + \lambda (\sum_{s=2}^S |\beta_s|) \right\}$

Recre
Subst

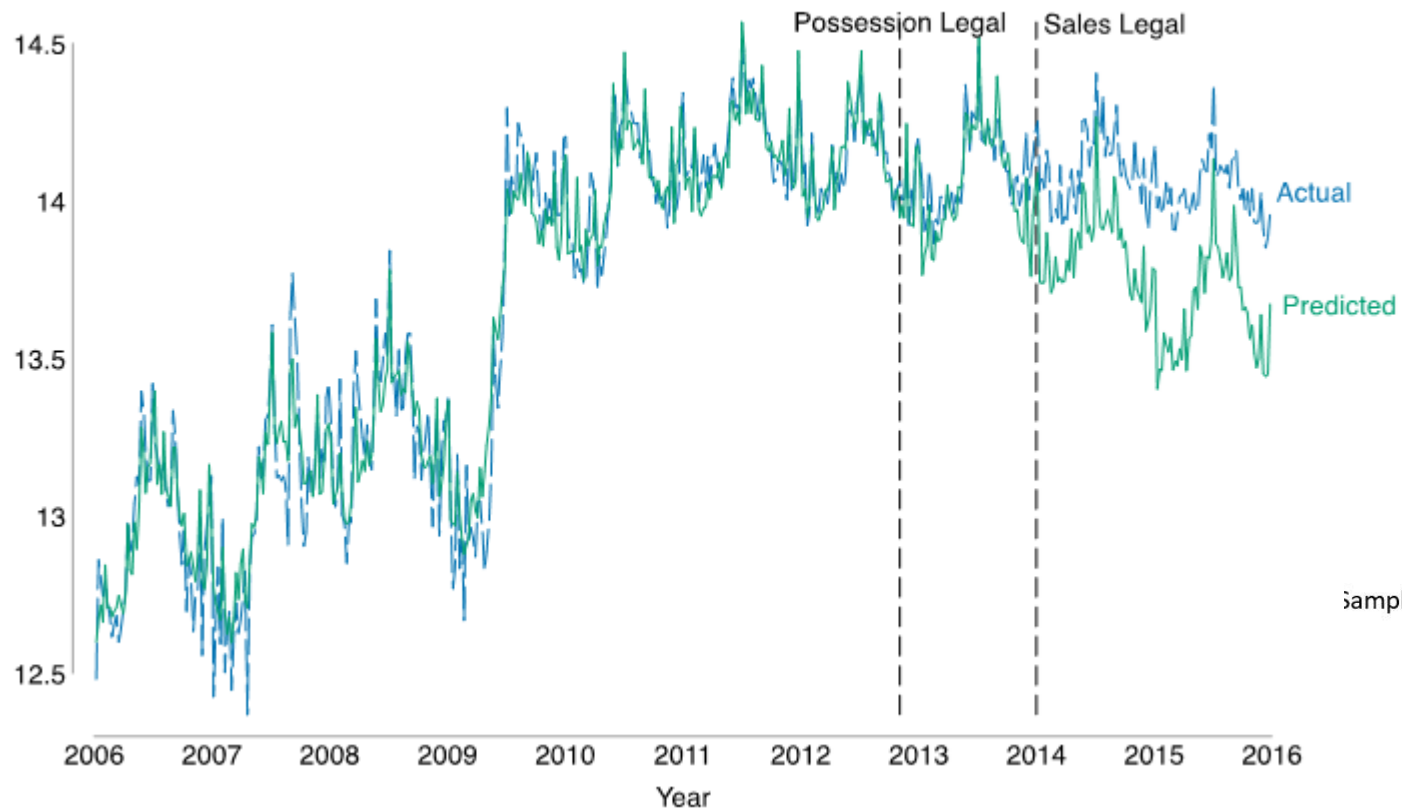
Monda
Salon v

Author

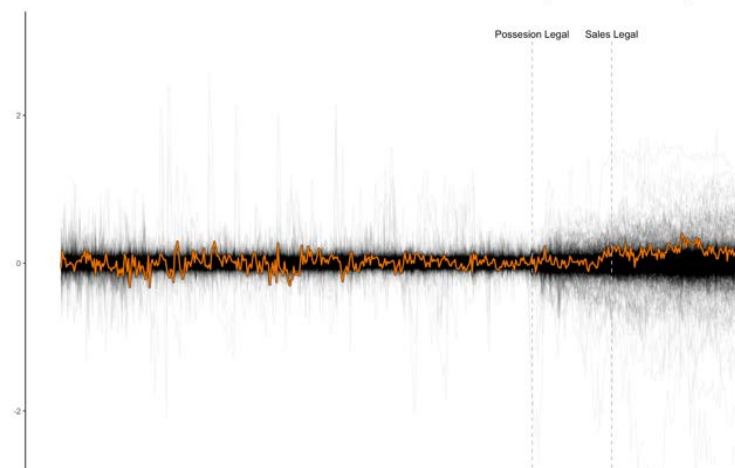
- The second term is the Lasso penalty function.
 - λ is a parameter that controls the penalty.
 - When $\lambda = 0$ you have OLS.
 - When $\lambda > 0$ you shrink the coefficients towards zero and sometimes you set some coefficients to zero. (Sparsity)
- Choose λ using cross-validation (Split the pre-period into training and evaluation sets)

Actual and Predicted Large Light Beer Sales in Colorado, 2006-2015

Inverse Hyperbolic Sine



Sampling Distribution: How well does the method work in placebo samples?



Example Student Group Interactions

SICE and SPEA at IndyBigData Yesterday



Females are a clear majority in high-growth industries and associated educational pipelines.

Gender imbalance is unsustainable in the current and projected economic and workforce environment.

Indiana University Bloomington

School of Informatics, Computing and Engineering

Vatsal Jatakia

Anurag Joshi

Roshith Raghavan

Olga Scrivner

Katy Borner

School of Public and Environmental Affairs

Livia Crim

David Lebhar

Kosali Simon

Coady Wing

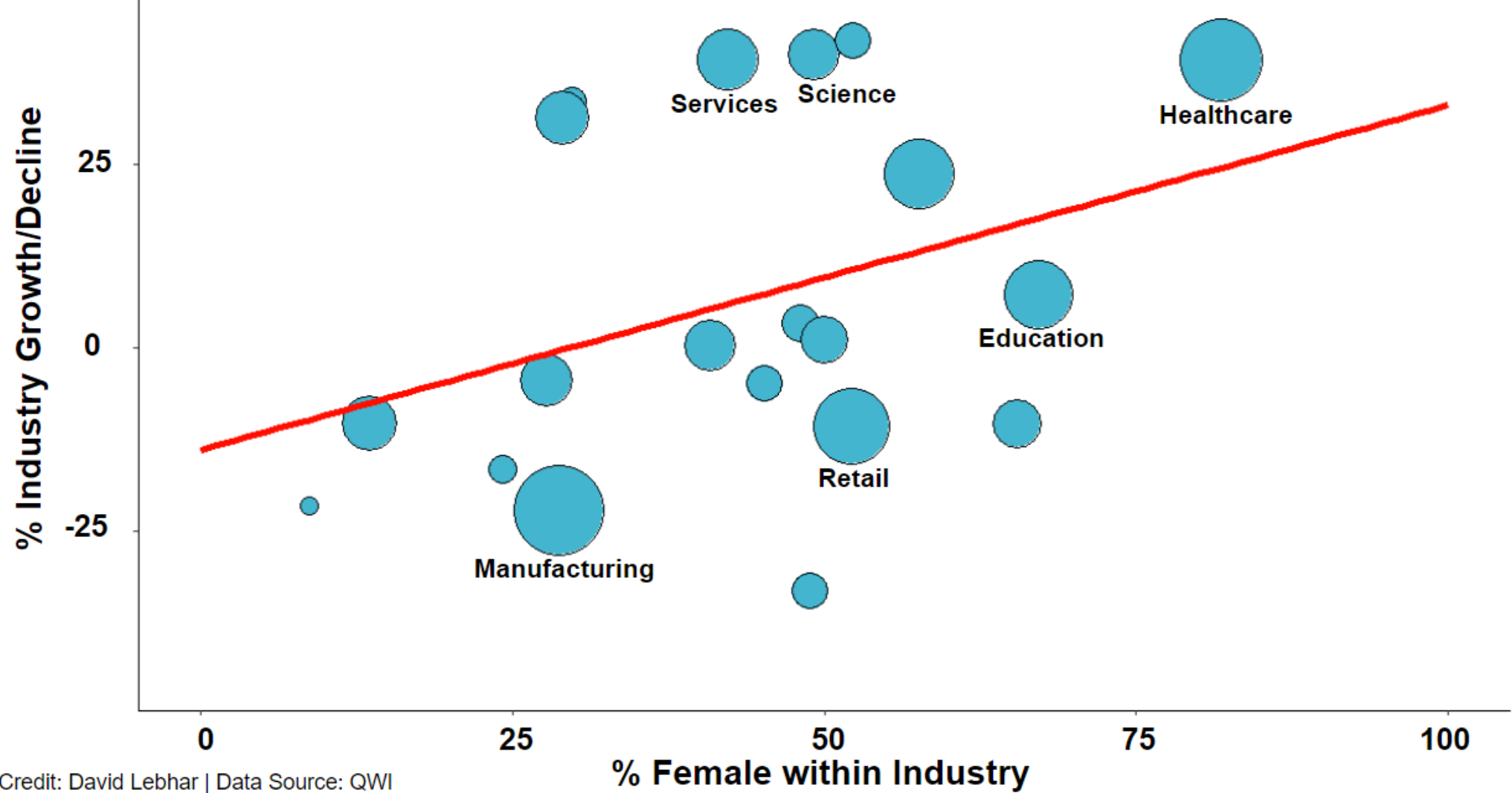


INDIANA UNIVERSITY
SCHOOL OF
INFORMATICS, COMPUTING,
AND ENGINEERING



SCHOOL OF
PUBLIC AND
ENVIRONMENTAL AFFAIRS
Indiana University

Correlation Between Industry Growth and Percent Female



Credit: David Lebar | Data Source: QWI

Forming Research Teams with DS and SS Expertise

Summer Research Kickoff Day

Health Policy Workshop

5/15/2018



Picture taken by Jim Hanchett, 5/15/2018

Summary

Traditionally developed separately

Many exciting new possibilities

Opportunities for Training Integration & Team Science