

The Journal of

Economic Perspectives

*A journal of the
American Economic Association*

Winter 2019

The Journal of Economic Perspectives

A journal of the American Economic Association

Editor

Enrico Moretti, University of California at Berkeley

Coeditors

Gordon Hanson, University of California at San Diego

Heidi Williams, Massachusetts Institute of Technology

Associate Editors

Leah Boustan, Princeton University

Dora Costa, University of California at Los Angeles

Janice Eberly, Northwestern University

David Figlio, Northwestern University

Amy Finkelstein, Massachusetts Institute of Technology

Fiona Scott Morton, Yale University

Betsy Stevenson, University of Michigan

Ebonya Washington, Yale University

Catherine Wolfram, University of California

Luigi Zingales, University of Chicago

Managing Editor

Timothy Taylor

Assistant Editor

Ann Norman

Editorial offices:

Journal of Economic Perspectives
American Economic Association Publications
2403 Sidney St., #260
Pittsburgh, PA 15203
email: jep@jepjournal.org

The *Journal of Economic Perspectives* gratefully acknowledges the support of Macalester College. Registered in the US Patent and Trademark Office (®).

Copyright © 2019 by the American Economic Association; All Rights Reserved.

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.

Printed by LSC Communications, Owensville, Missouri, 65066, USA.

No responsibility for the views expressed by the authors in this journal is assumed by the editors or by the American Economic Association.

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Winter 2019, Vol. 33, No. 1. The *JEP* is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. Annual dues for regular membership are \$22.00, \$32.00, or \$42.00 depending on income; for an additional \$15.00, you can receive this journal in print. The journal is freely available online. For details and further information on the AEA go to <https://www.aeaweb.org/>. Periodicals postage paid at Nashville, TN, and at additional mailing offices.

POSTMASTER: Send address changes to the *Journal of Economic Perspectives*, 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the U.S.A.

The Journal of
Economic Perspectives

Contents

Volume 33 • Number 1 • Winter 2019

Symposia

Women in Economics

- Shelly Lundberg and Jenna Stearns, “Women in Economics: Stalled Progress” 3
Leah Boustan and Andrew Langan, “Variation in Women’s Success across PhD
Programs in Economics” 23
Kasey Buckles, “Fixing the Leaky Pipeline: Strategies for Making Economics Work
for Women at Every Stage” 43

Financial Stability Regulation

- Daniel K. Tarullo, “Financial Regulation: Still Unsettled a Decade After the
Crisis” 61
Darrell Duffie, “Prone to Fail: The Pre-Crisis Financial System” 81
David Aikman, Jonathan Bridges, Anil Kashyap, and Caspar Siegert, “Would
Macroprudential Regulation Have Prevented the Last Crisis?” 107

Public Provision of Economic Data

- Ellen Hughes-Cromwick and Julia Coronado, “The Value of US Government
Data to US Business Decisions” 131
Hugh Rockoff, “On the Controversies Behind the Origins of the Federal
Economic Statistics” 147
Ron S. Jarmin, “Evolving Measurement for an Evolving Economy: Thoughts on
21st Century US Economic Statistics” 165

Articles

- Spencer Banzhaf, Lala Ma, and Christopher Timmins, “Environmental Justice:
The Economics of Race, Place, and Pollution” 185
Susan Athey and Michael Luca, “Economists (and Economics) in Tech
Companies” 209
Ariel Pakes and Joel Sobel, “Parag Pathak: Winner of the 2018 Clark Medal” 231

Features

- Timothy Taylor, “Recommendations for Further Reading” 247

Statement of Purpose

The *Journal of Economic Perspectives* attempts to fill a gap between the general interest press and most other academic economics journals. The journal aims to publish articles that will serve several goals: to synthesize and integrate lessons learned from active lines of economic research; to provide economic analysis of public policy issues; to encourage cross-fertilization of ideas among the fields of economics; to offer readers an accessible source for state-of-the-art economic thinking; to suggest directions for future research; to provide insights and readings for classroom use; and to address issues relating to the economics profession. Articles appearing in the journal are normally solicited by the editors and associate editors. Proposals for topics and authors should be directed to the journal office, at the address inside the front cover.

Policy on Data Availability

It is the policy of the *Journal of Economic Perspectives* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the above requirements cannot be met.

Policy on Disclosure

Authors of articles appearing in the *Journal of Economic Perspectives* are expected to disclose any potential conflicts of interest that may arise from their consulting activities, financial interests, or other nonacademic activities.

Journal of Economic Perspectives Advisory Board

Stephanie Aaronson, Brookings Institution

Janet Currie, Princeton University

Claudia Goldin, Harvard University

Robert E. Hall, Stanford University

Kenneth Kuttner, Williams College

Trevon Logan, Ohio State University

Eduardo Porter, *New York Times*

Paul Romer, New York University

David Sappington, University of Florida

Gianluca Violante, Princeton University

Ludger Woessmann, Ifo Institute for Economic Research

Women in Economics: Stalled Progress

Shelly Lundberg and Jenna Stearns

Although women are still a minority in the economics profession, female representation in the discipline has increased slowly over the past century. By the mid-2000s, just under 35 percent of PhD students and 30 percent of assistant professors were female, but these numbers have remained roughly constant ever since. This is not the first time progress on the path to gender equality in economics has stalled: women were more prominent as researchers in the early years of the 20th century than they were mid-century. Listings of dissertations in progress in the *American Economic Review* show that women were writing 6 percent of US PhD dissertations in 1912, rising to a peak of nearly 20 percent in 1920 but then falling back to 7 percent by 1940 (Forget 2011). Forget (2011) links the decline in female representation in academic economics to the emergence of home economics and social work as academic fields, the expansion of employment opportunities in government, and increased hostility and overt discrimination in economics departments. Cherrier (2017b) draws a parallel between these trends in economics and the defeminization of computer science as this field became increasingly professionalized, “scientized,” and lucrative after the mid-1980s.

■ *Shelly Lundberg is the Leonard Broom Professor of Demography, University of California Santa Barbara, Santa Barbara, California. She is also a Research Fellow, Institute for the Study of Labor (IZA), Bonn, Germany, and Adjunct Professor of Economics, University of Bergen, Bergen, Norway. Jenna Stearns is an Assistant Professor of Economics, University of California Davis, Davis, California. Lundberg is the corresponding author at slundberg@ucsb.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.33.1.3>

doi=10.1257/jep.33.1.3

Common explanations for women's underrepresentation in economics in the mid-20th century included comparative advantage and diverging preferences by gender. By the early 1970s, however, overt discrimination was blamed for "the bizarre and irrational underrepresentation of women in the economics profession" (CSWEP 1973). The contested establishment of the Committee on the Status of Women in the Economics Profession (CSWEP) at the 1971 American Economic Association (AEA) business meeting took place in the wake of public discussion and government action on discrimination, actions by other professional associations to increase the representation of women, and growing interest in discrimination as an economic phenomenon with the early work by Becker and Arrow (Cherrier, Chassonnery-Zaigouche, and Singleton 2018). A Caucus of Women Economists drafted resolutions requiring the AEA to adopt "a positive program to eliminate sex discrimination." The resolutions were presented at the business meeting, where they provoked heated debate and several speeches in opposition, but were approved by a vote of the attending membership. The room had been packed by progressive economists prior to the vote, according to a first-hand account by Strober (2016, chap. 6). Though the beginning statement "Resolved that the American Economic Association declares that economics is not a man's field," was amended to insert "not exclusively," the resolutions were adopted in full, including the establishment of CSWEP (Cherrier 2017b). Feminist activism scored similar successes throughout the academy during this period: women's committees were established in the American Sociological Association in 1970, and in the American Physical Society in 1972.

Despite large gains in female representation in economics in the 1970s and 1980s, reactions to women's progress were mixed. In the Fall 1998 issue of this journal, the 25th anniversary of CSWEP was commemorated with a symposium that reflected contrasting views of efforts to diversify the profession. The various contributions reviewed women's progress in economics favorably but expressed concern about the implications of low representation of women among economics undergraduates (Bartlett 1998); critiqued CSWEP's nonmilitancy and offered unfavorable comparisons with more activist women's committees in other fields (Bergmann 1998); and asserted that, partly as a result of CSWEP's activities, the "pendulum has probably swung too far so that men are the ones currently being discriminated against" (Friedman 1998).

Since then, women's progress in academic economics has slowed, with virtually no improvement in the female share of junior faculty or graduate students in decades. Little consensus has emerged as to why, though there has been a renewal of widespread interest in the status and future of women in economics and of the barriers they face to professional success. In this paper, we first document trends in the gender composition of academic economists over the past 25 years, the extent to which these trends encompass the most elite departments, and how women's representation across fields of study within economics has changed. We then review the recent literature on other dimensions of women's relative position in the discipline, including research productivity and income, and assess evidence on the

barriers that female economists face in publishing, promotion, and tenure. While differences in preferences and constraints may directly affect the relative productivity of men and women, productivity gaps do not fully explain the gender disparity in promotion rates in economics. Furthermore, the progress of women has stalled relative to that in other disciplines in the past two decades. We propose that differential assessment of men and women is one important factor in explaining this stalled progress, reflected in gendered institutional policies and apparent implicit bias in promotion and tenure processes.

Women in PhD-granting Economics Departments, 1972–2017

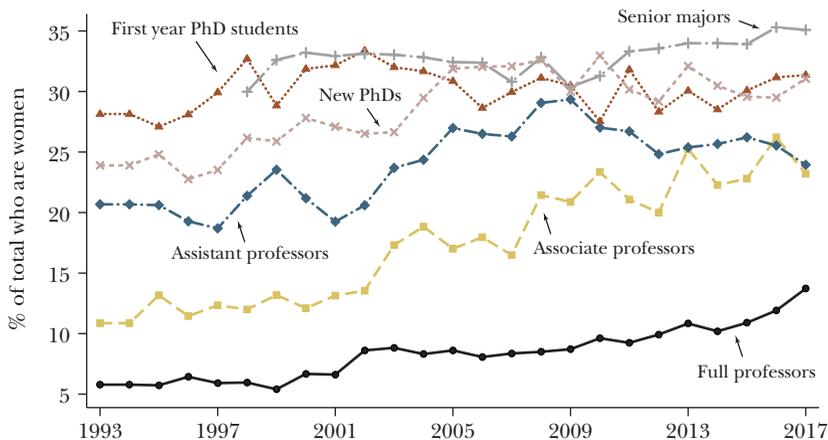
In 1972 and 1973, the Committee on the Status of Women in the Economics Profession conducted surveys of economics departments “to remedy the total lack of information on how many women economists exist or are currently being trained” (Bell 1973). This task was assumed by the American Economic Association in 1974, and questions about faculty and graduate student gender were combined with other data requests to form the Universal Academic Questionnaire (UAQ) sent to academic departments. The UAQ provided the data for CSWEP’s reports on the status of women in economics until 1993, when the committee resumed their separate survey to improve response rates. This survey gathers information each fall on the gender composition of new and graduating PhD students; faculty at the assistant, associate, and full professor levels; nontenure track faculty; and senior undergraduate majors. Most of the data presented in this section combines the CSWEP and UAQ data for PhD-granting departments from 1993 to 2017.¹

We can provide a longer-term perspective for one important set of departments. The microdata from the first CSWEP survey has been lost, but the 1972 CSWEP Annual Report, published in the *AER Papers and Proceedings* issue, includes aggregate results for one identifiable group of 43 departments—the Chairman’s Group. This group was known informally as “the cartel,” because the chairs met every year for breakfast at the ASSA meetings and discussed planned salary offers for new assistant professors. These departments are listed in the report, and with the exception of the University of Rochester, all of them responded to the initial survey.² The departments in the Chairman’s Group granted about two-thirds of US economics PhDs in the early 1970s, and we can track the faculty and graduate student gender composition in this set of highly ranked departments over a 45-year period.

¹ Response rates to the CSWEP survey of PhD-granting departments have been 100 percent in recent years, but below that prior to 2015—nonresponses are replaced by UAQ data when possible. The data for 2000 has been lost. The cleaned data were produced by the Inter-university Consortium for Political and Social Research (ICPSR) under the direction of Margaret Levenstein, and is available to researchers through ICPSR. About 4 percent of the observations are imputed.

²The report also includes aggregate gender ratios for “all departments” based on 397 questionnaires returned out of 1364 questionnaires sent (Bell 1973).

Figure 1

Representation of Women among First-Year PhD Students, New PhDs, and Faculty by Rank for the Chairman's Group of Departments, 1993–2017


Source: Authors, using data from CSWEP and from the UAQ for PhD-granting departments from 1993 to 2017.

Note: The Chairman's Group consists of Brown University, University of California—Berkeley, University of California—Davis, University of California—Los Angeles, Carnegie Mellon University, University of Chicago, University of Colorado, Columbia University, Cornell University, Duke University, University of Florida, Harvard University, University of Illinois, Indiana University, Iowa State University, Johns Hopkins University, University of Maryland, Massachusetts Institute of Technology, University of Michigan, Michigan State University, University of Minnesota, New York University, State University of New York—Buffalo, University of North Carolina—Chapel Hill, Northwestern University, Ohio State University, University of Pennsylvania, University of Pittsburgh, Princeton University, Purdue University, University of Rochester, University of Southern California, Stanford University, Texas A & M University, University of Texas—Austin, Vanderbilt University, University of Virginia, University of Washington—Seattle, Washington State University, Washington University in St. Louis, Wayne State University, University of Wisconsin, and Yale University.

Substantial progress was made during the 1970s and 1980s in the representation of female faculty within the Chairman's Group departments. In 1972, women accounted for only 2 percent of full professors, 4 percent of associate professors, and 9 percent of assistant professors. By the time the CSWEP survey was resumed in 1993, the fraction of full professors who were female had tripled to 6 percent, 11 percent of associate professors were women, and the female share of assistant professors had more than doubled to 21 percent.

Figure 1 shows that the proportion of senior female faculty in the Chairman's Group continued to grow slowly from 1993 to the present. Among full professors, the female share increased from 6 percent to more than 13 percent, and among associate professors, from 11 to 23 percent. For assistant professors, however, the pattern is somewhat different: the share of women increased from 20 percent in 1993 to 29 percent in 2009, and then decreased over the past decade to 24 percent, leaving

little net growth at junior ranks over the past 24 years.³ Over the same period, there has also been little improvement in female representation among first-year PhD students, from 28 percent in 1993 to an average of 30 percent in the past five years. (During the 1990s, there was a consistent gap of a couple of percentage points between the female share of first-year graduate students and exiting PhDs five years later that seems to indicate higher attrition for female graduate students, but this gap disappeared by the entering class of 2000.) This stasis extends to undergraduate study of economics as well: the female share of senior economics majors has remained between 30 and 35 percent since the data series began in 1998. Progress towards gender equality at the intake levels of the profession appears to have ceased (with some deterioration for junior faculty), while women's representation at senior levels continues to rise, fueled for now by the entry of women into academic economics in past decades.

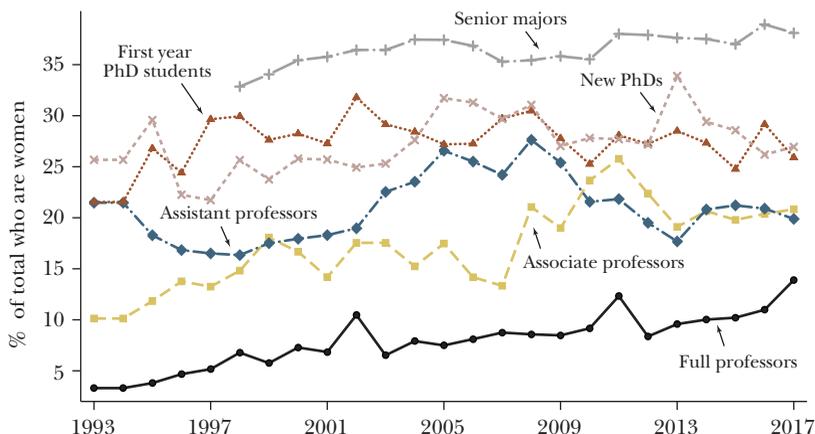
Although the Chairman's Group does not provide a complete picture of PhD departments, there are reasons to be particularly interested in the progress that women have made in elite departments. Economics is a very hierarchical social science (as discussed in this journal by Fourcade, Ollion, and Algan 2015), and a high fraction of both the articles published in top journals and the faculty who train PhD students come from the most highly-ranked departments. Figure 2 shows the 1993–2017 faculty and student data for the departments rated in the top 20 by US News and World Report. The data is a bit noisier for these smaller samples than for the Chairman's Group, but some trends are clear.

In top 20 programs, the representation of women among full professors was only 3 percent in 1993, grew slowly to 10 percent in recent years, and then rose to nearly 14 percent in 2017. The female fraction of associate professors (which grew steadily throughout this period in the Chairman's Group), increased from 10 percent to as high as 26 percent in 2011, but has declined in recent years to about 20 percent. Female representation among assistant professors stood at about 21 percent in 1993, reached a peak of 27.6 percent in 2008, and has since fallen back to 20 percent, meaning that no net progress has been made at the junior faculty level in top 20 departments over the past 24 years. These patterns are qualitatively similar if we look only at the top 10 programs in the US News and World Report Rankings as well.

To compare women's progress in economics to other academic disciplines, we have combined the data from the Committee on the Status of Women in the Economics Profession on the Chairman's Group with data on the share of female faculty by rank in top-50 departments for several science and social science disciplines. These data, for 2002, 2005, 2007, and 2012, come from the Nelson (2004) Diversity Surveys of department chairs, collected under the auspices of the University

³This decrease is not apparent in the data on all PhD-granting departments that is presented in the CSWEP annual report (CSWEP 2018), where the assistant professor gender ratio appears flat for the 2005–2017 period. A separate analysis confirms that, for the non-Chairman's Group departments (which tend to be lower-ranked than the Chairman's Group), female representation among assistant professors has continued to grow slowly.

Figure 2
Representation of Women among First-Year PhD Students, New PhDs, and Faculty by Rank: Top 20 Economics Departments, 1993–2017



Source: Authors, using data from CSWEP and from the UAQ for PhD-granting departments from 1993 to 2017.

Note: The departments included are Brown University, Carnegie Mellon University, Columbia University, Cornell University, Duke University, Harvard University, Massachusetts Institute of Technology, New York University, Northwestern University, Princeton University, Stanford University, University of California—Berkeley, University of California—Los Angeles, University of California—San Diego, University of Chicago, University of Michigan—Ann Arbor, University of Minnesota, University of Pennsylvania, University of Wisconsin—Madison, and Yale University.

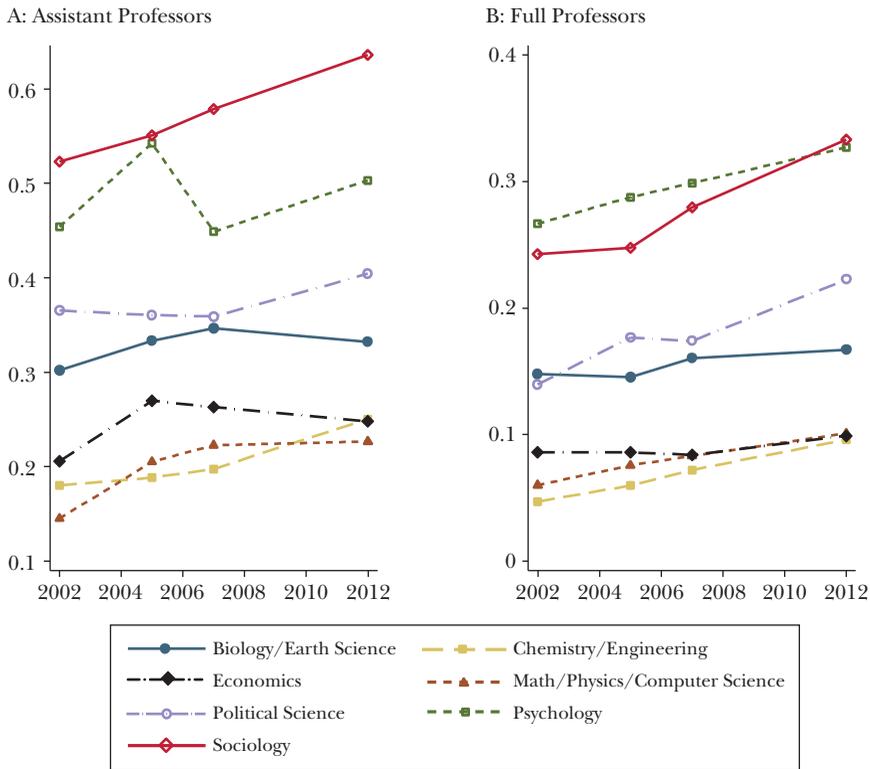
of Oklahoma. Top-50 departments are as ranked by the National Science Foundation according to field-specific research expenditures.⁴ Figure 3 shows trends in the share of female assistant and full professors across seven discipline groups. For ease of presentation, we combine data from chemistry and four types of engineering departments (chemical, civil, electrical, and mechanical). We also combine biology and earth science, and math, computer science, physics, and astrophysics. In general, there is an upward trend in the share of female faculty at all ranks over this ten-year period. Hard sciences have the lowest share of female professors at all ranks, while the social sciences have the highest. Economics remains solidly within the lowest group in terms of female faculty shares, alongside physics, math, and engineering, and far below the biological and other social sciences. At the senior level, economics seems to have lost some ground relative to other sciences during

⁴Comparable data on top-50 departments is not available going back further in time. Using the NSF Survey of Doctoral Recipients (SDR), Ginther and Kahn (2004) and Ceci, Ginther, Kahn, and Williams (2014) show trends in the share of female assistant professors and tenured faculty across disciplines since 1973. However, the SDR samples doctoral recipients from all US academic institutions, and is not necessarily representative of faculty at top departments.

Figure 3

Representation of Women in Top-50 Departments, 2002–2012

(share female)



Source: Authors, using data from the Nelson Diversity Surveys and CSWEP.

this period. In all cases the share of women is decreasing with rank (note the y-axes for graphs A and B are different).

Do Women Study Different Fields of Economics than Men and Has the Distribution of Women across Fields Changed over Time?

While the survey data from the Committee on the Status of Women in the Economics Profession allow us to track the career progression of female academic economists over time, much less is known about another dimension of women’s representation in economics—their distribution across fields of study. Understanding how trends in research areas differ for men and women may be important for understanding differential trends in publishing and tenure. Field choice may

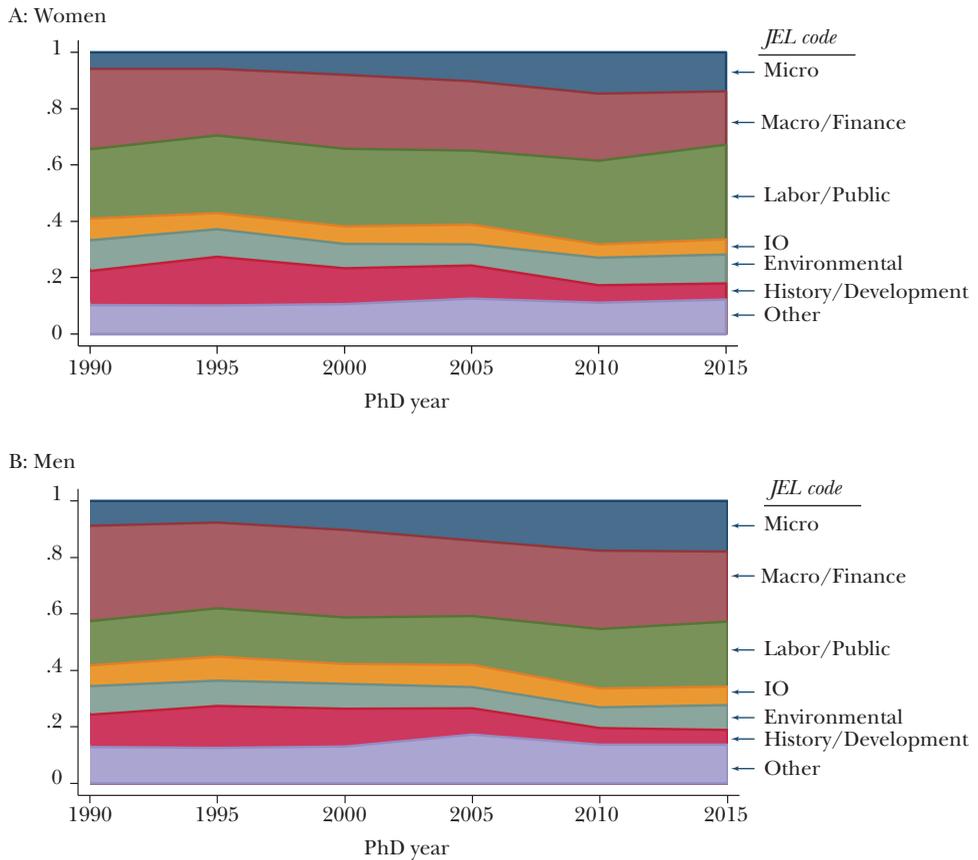
affect entry into tenure-track positions in economics, publication rates, and the probability of publishing in top journals, all of which may also affect the probability of earning tenure. However, the limited evidence estimating differences in professional success across economics fields is mixed. Recent work shows that field choice explains a large share of the gender gap in research output (Ductor, Goyal, and Prummer 2018), while Ginther and Kahn (2004) find that broad fields are an insignificant predictor of tenure among a sample of assistant professors in 1989. Several recent papers in economics document the contemporary distribution of women across fields, but to our knowledge, the existing research cannot provide insight into how fields of study have changed over the past few decades.

Using data from the National Bureau of Economic Research Summer Institute 2001–2016, Chari and Goldsmith-Pinkham (2017) show that the distribution of female economists at this event is not uniform across fields. Women are particularly scarce in macro and finance, and more abundant in labor and other applied microeconomic fields. Beneito, Pilar, Boscá, Ferri, and García (2018) use data from the annual AEA meetings from 2010–2016 to show the percentage of female authors in five subfields according to the *Journal of Economic Literature* subject codes of the sessions. For the most recent years, the authors also use machine learning to classify the paper abstracts by topic. Similar to Chari and Goldsmith-Pinkham (2017), they find that female representation is substantially lower in macro, finance, and mathematical and quantitative methods than in applied micro and other fields. An important caveat about these findings is that participation at both the AEA annual meetings and the NBER Summer Institute may be nonrepresentative across both gender and field, and again, little is known about how the gender composition across economic fields has changed over a longer period.

To provide a broader perspective on the evolution of women across fields and over time, we have collected information on recipients of PhDs in economics from 1991–2017, including the recipient's name and the JEL code of their dissertation. This information comes from the Doctoral Dissertations in Economics lists published annually in the *Journal of Economic Literature*, and represents almost all major PhD-granting departments in the United States. To classify the gender of each doctoral recipient, we use two databases that allow us to determine the probability that a given name is female: the Social Security Administration name files and the Genderize.io database for an international dimension. We match the first names in our data to these probabilities, and assign gender to those with a probability of being female that is above 0.8 or below 0.2.⁵ In total, we identify the gender

⁵The first database is the Social Security Administration name files, which include all names with at least five occurrences in a given year based on applications for a US Social Security card at birth. Because this data is only representative of US-born individuals, and a large share of PhD recipients in economics are foreign-born, we also match to the Genderize.io database, which contains over 200,000 distinct names from 79 countries. Both datasets contain the number of male and female incidences of the name. We designate a name as female if the probability that the name is female is higher than 0.8, and male if the probability is lower than 0.2. We are able to match 88.5 percent of the individuals in our data to a name in at least one of the two databases, and we assign a gender to 83 percent of the total sample.

Figure 4
Dissertation Topics of Women and Men by Year
 (share)



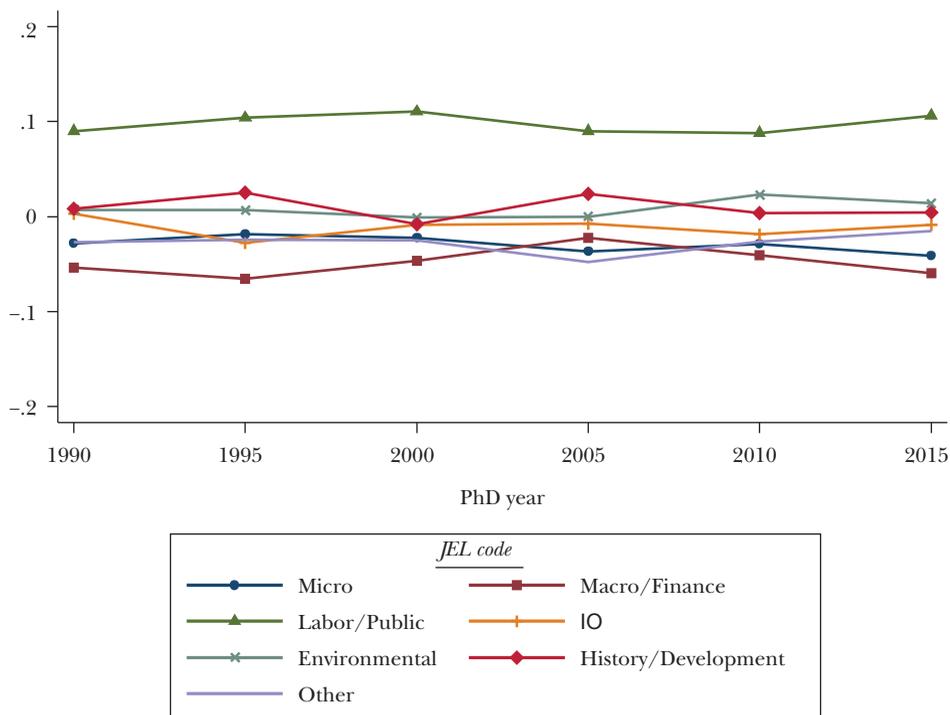
Source: Authors, using data from the annual list of Doctoral Dissertations in Economics, 1991–2017.
 Note: Data collapsed into five-year bins for smoothness. The 1990 bin contains data from 1991–1994 and the 2015 bin contains data from 2015–2017; all other bins contain 5 years of data.

of 23,442 out of 28,209 individuals over 26 years. About 29 percent of new PhDs over this period are female, and the trend in the share of female PhD recipients over time mirrors the CSWEP data above.

Figures 4A and 4B show the distribution of women and men across seven fields over time.⁶ In general, the distributions of men and women across these fields are very similar. The higher representation of women in labor/public is apparent,

⁶We have collapsed the JEL codes into seven categories for ease of presenting results. “Micro” is JEL code D; “Macro/Finance” is codes E, F, and G; “Labor/Public” is H, I, and J; “IO” is L; “Environmental” is Q; “History/Development” is N and O; and “Other” contains the remaining JEL codes A, B, C, K, M, P, R, Y, and Z, which all represent a relatively small share of PhD dissertations.

Figure 5

Difference between Share of Women and Share of Men in Particular Fields of Economics

Source: Authors, using data from the annual list of Doctoral Dissertations in Economics, 1991–2017.

Note: Data was collapsed into five-year bins for smoothness. The 1990 bin contains data from 1991 to 1994 and the 2015 bin contains data from 2015 to 2017; all other bins contain five years of data.

but female economists are well-represented in all fields. In terms of changes over time, women are more likely to write dissertations in micro and labor/public than they were in the early 1990s, and somewhat less likely to study macro/finance and history/development. However, in large part these trends reflect broader trends in the profession, and very similar changes in field choice can be seen among men.

In fact, the distribution of PhD recipients across fields has not evolved differentially for men and women since the early 1990s. To show this more clearly, Figure 5 plots the difference between the share of women in a particular field and the share of men in that field over time. While it is certainly the case that women are more likely than men to study topics in labor and public economics and less likely to do dissertation research in macro and finance across the entire time period, there is virtually no evidence of differential trends. (Because of the gender imbalance in economics, there are still more men than women who graduate with a dissertation classified as labor or public every year.) It is not entirely clear why a higher fraction of women than men choose labor-oriented research topics. One commonly

discussed hypothesis is that women on average have stronger interests in studying individual behavior. A survey of AEA members in 2008 showed that, while there were no gender differences in responses to beliefs about core economic concepts, female economists are more likely to support the need for government intervention versus market solutions (May, McGarvey, and Whaples 2014). This bias in choice of field could be sustained over time if the research environment across different fields is an important factor in what graduate students choose to study; that is, the higher share of female faculty in labor economics might encourage female students to study labor through role model effects.

This lack of change in the relative gender composition across fields over time is important for two reasons. First, it suggests that as the share of female PhD recipients has risen, the more recent female cohorts are no different in terms of their broad research interests. Second, differential trends in field choice over time cannot explain the observed changes in the gender gap in the share of PhD recipients who become assistant professors and who are later tenured.

The graphs shown here use data starting in 1991 because this was the last time that the *Journal of Economic Literature* subject codes were substantially redesigned. Focusing on this period means that the JEL codes are comparable over time. However, it is possible to collect similar data going back further. For a longer-term perspective, we have also compiled data from the early 1970s. JEL codes in this period were not completely comparable with those used today. In particular, “core” areas of economic theory including micro and macro theory were categorized together in a “General Economics” category, though the applied categories are reasonably comparable for our purposes.⁷ There is still no evidence of differential trends by gender in these applied areas. In particular, the difference in the share of women compared to men who study labor and public economics has remained constant at about 0.1 since at least the early 1970s.

How Do Women’s Academic Careers in Economics Compare With Men’s?

Women’s representation in economics departments tends to fall as academic rank increases. As shown above (Figures 1 and 2), the female share of full professors in research-oriented departments ranges from 8 to 13 percent, from 20 to 25 percent for assistant professors, and from about 25 to 30 percent for PhD students. Simple “lock-step” models tracking cohorts of PhD recipients, reported annually by the Committee on the Status of Women in the Economics Profession, show a distinct drop-off from last-year-in-rank assistant professors to last-year-in-rank

⁷The full category name is “General Economics; including Economic Theory, History of Thought, Methodology, Economic History, and Economic Systems.” The comparable categories include environmental, development, IO, labor/public, and international economics. See Cherrier (2017a) for a history of JEL codes, including a list of categories in this time period.

associate professors for PhD cohorts from the mid-1980s through 2003 (CSWEP 2018). This suggests that the economics career pipeline is “leaky” at the stage when most academics receive tenure.

Studies using micro-data tend to confirm that something goes wrong for female economists at the tenure stage. Using longitudinal data on all AEA members from the 1960s through the 1980s, McDowell, Singell, and Ziliak (2001) find that women were less likely than men with similar characteristics to be promoted to both associate professor and full professor. However, they also find that women’s promotion prospects improved in the 1980s, leaving no unexplained gender differences in promotion for individuals observed in 1989. In contrast, Ginther and Kahn (in this journal, 2004) find clear evidence of a leaky pipeline in a sample restricted to AEA members who were assistant professors at PhD-granting departments in 1989—women in this sample were less likely to get tenure than men and took longer to achieve it. Ten years after receiving their PhDs, female economists were 21 percentage points less likely than men to have a tenured academic job. Differences in productivity, including number of publications, publication quality, and citations, explained only 30 percent of this promotion gap. In the same paper, Ginther and Kahn find a similar result using the 1972 to 1991 PhD cohorts from the National Science Foundation Survey of Doctoral Recipients, which has limited data on publication quality but does have information on family characteristics. Controlling for the presence of young children, which had an impact on promotion independent of productivity, leaves a substantial portion of the gender difference in tenure probabilities unexplained. The authors conduct the same analysis for other disciplines using the Survey of Doctoral Recipients data, and found that the gender promotion gap in economics was distinctive. There were negligible gender gaps in the transition to tenure in statistics and the sciences, and only an 8 percent gap in the other social sciences. In engineering, women were more likely than men to have been promoted after ten years.

In a later study of women’s careers in academic social science that examined cohort differences using the 1981–2008 waves of the Survey of Doctoral Recipients, Ginther and Kahn (2014) find that, although there were gender differences in tenure probabilities for the 1980 cohort of PhDs in other social science disciplines, these had disappeared for the 1999 PhDs, while a 20 percent gender gap persisted in economics. They conclude: “Economics is the one field where gender differences in tenure receipt seem to remain even after background and productivity controls are factored in and even for single childless women” (p. 311). Similarly, they find no significant gender differences in promotion to tenure or full professor in the sciences overall after controlling for demographic, family, and productivity covariates (Ginther and Kahn 2009).

In an omnibus study on women in academic science written in collaboration with two psychologists, Ginther and Kahn examine recent career progression in math-intensive fields of study and find evidence of gender inequality only in economics (Ceci, Ginther, Kahn, and Williams 2014). Most of these disciplines made progress towards gender equality in income and promotion between the mid-1990s

and 2010, while economics did not. In geosciences, engineering, math/computer science, and physical sciences, men and women now enter PhD programs at rates proportionate to their representation in college majors and are equally likely as PhD students to be hired into tenure-track positions.

In addition to the persistent gender gap in promotions to tenured positions, Ceci et al. (2014) also find significant gaps in academic salaries and job satisfaction among economists that have not decreased (and in some cases have increased) over time. In the 1995 Survey of Doctoral Recipients data, female assistant professors in economics were paid lower salaries than male assistant professors, but the difference was not significant. By 2010, the average salary gap in these data had increased and become significant. Over the same period, the relative salaries of female full professors fell as well, to 74 percent of male salaries by 2010, though there are no significant salary differences at research-intensive universities with PhD programs (so-called “R1” institutions). Women in the sciences tend to report being less satisfied with their jobs than male scientists, but the gender differences in the 1997 and 2010 Survey of Doctoral Recipients were generally small and fell over time. The gap in job satisfaction among economists, in contrast, was large in 1997 and grew by 2010, with men becoming more likely to report being very or somewhat satisfied with their jobs and women becoming less likely to do so.

Gender gaps in job satisfaction may not be surprising, given the disadvantages women appear to face in promotion and pay, but what might explain these differences in substantive career progression? Many studies have shown that women in science, technology, engineering, and mathematics fields, including economics, have fewer publications than men at equivalent stages of their career, though there appears to be no difference in hours worked (Ceci et al. 2014). Ginther and Kahn (2004) report that, ten years post-PhD, women in the 1989 cohort of assistant professors have 0.3 fewer top-10 publications and 3.8 fewer articles in other journals, though these differences do not explain most of the promotion gap. The 1995 and 2008 Survey of Doctoral Recipients data includes the number of articles accepted in refereed journals in the previous five years and, according to this metric as well, female assistant professors published less than male assistant professors. Between 1995 and 2008, this gap increased and became significant, with women publishing less and men publishing more (Ceci et al. 2014). A recent study based on a broader database of journal articles from EconLit (with gender identified for 80 percent of authors) finds that the raw gender gap in research output for all economists has been relatively constant at around 50 percent since the late 1980s, though 43 percent of this gap can be explained by differences in experience and field (Ductor, Goyal, and Prummer 2018).

A leading hypothesis for why female academics are less productive is that women have more intense domestic responsibilities; indeed, the evidence from most science, engineering, technology, and mathematics fields is that publications by single childless females are not significantly different from publications by single childless men. This is not the case in economics and the physical sciences, however, where there is a significant gender gap among the childless as well. Gender norms

that assign more nurturing roles to women may also influence productivity through the way that time on the job is allocated. Studies of faculty in science, technology, engineering, and mathematics have found a gender discrepancy in time use, with women spending more time on teaching, service, and other nonresearch academic activities such as mentoring students (Xie and Shauman 2003; Misra, Lundquist, Holmes, and Agiomavritis 2011). Female faculty are more likely to volunteer for low-reward tasks (tasks unlikely to contribute to one's chances for promotion), and lab experiments confirm that women volunteer, and are asked to volunteer, more than men (Babcock, Recalde, Vesterlund, and Weingart 2017). However, we are not aware of any economics-specific evidence on professional time allocation.

Evidence for Barriers

If women's relative failure to advance in departments of economics cannot be explained by the gender gap in productivity, the possibility of differential treatment arises. As we will discuss, a number of recent papers explore the role of gender per se in the economics profession, examining the possible causes of differential attrition and the persistent gap in tenure probabilities. Taken together, this work builds a case that female economists face substantial barriers throughout their career. These barriers may influence persistence in the profession by reducing expectations of future success, impeding research activity and publication outcomes, or affecting the probability of promotion even conditional on observed productivity.

Barriers that act to limit women from becoming tenured economists may start earlier. For example, Figure 1 shows that only about one-third of undergraduate economics majors are women. Also, Figure 1 shows that attrition rates in economics PhD programs were higher for women than men until the mid-2000s (as shown by the gap between the share of women who were first-year PhD students and the share that were new PhDs). In this symposium, the paper by Buckles discusses the research on policies that have been used in trying to raise the share of women at all stages of the economics career pipeline, while the paper by Boustan and Langan looks at the heterogeneity across departments in the share of women admitted to and completing PhD programs. We focus here primarily on issues affecting the research productivity of female economists.

An accumulating body of evidence suggests that early-career female economists may be adversely affected by limited access to the mentoring and social networks that support research activities, as well as by potential biases in the referee process. For example, a lack of senior female mentors may disadvantage assistant professors, especially if important information about publishing and tenure is transmitted informally within departments or research networks. In an effort to expose female assistant professors to successful female role models, boost research productivity, and help prepare them for the tenure process, the CSWEP Mentoring Program, CeMENT, matches junior female faculty with senior mentors. The program has been routinely oversubscribed, enabling a randomized control trial of the program

to be conducted in the 2000s. This evaluation found that CeMENT significantly increased the publication rates and grant funding of participants, bolstering the argument that a lack of mentoring may be important for women (Blau, Currie, Croson, and Ginther 2010).

Barriers in social network formation that hinder mentoring in a male-dominated field may lead men and women to have different research collaboration and coauthorship networks as well (McDowell, Singell, and Stater 2006). Although women in economics have a higher share of coauthored papers, their coauthorship patterns are distinct from those of men in ways that are predictive of lower output—fewer coauthors, higher clustering, and more collaboration with the same coauthors (Ductor, Goyal, and Prummer 2018). Coauthored publications also appear to be evaluated differently based on the gender of the authors. Male and female economists receive similar credit for sole-authored papers of similar quality in terms of their impact on tenure decisions (Sarsons 2017a). However, women receive significantly less credit for coauthored work, particularly when they coauthor with men. This contrasts with evidence from sociology, where Sarsons finds that men and women benefit equally from coauthored work.

Women and men in economics may also face different experiences throughout the publishing process. Several papers have tested for outright discrimination against women in manuscript review, but the empirical evidence is mixed. Ferber and Teiman (1980) study double-blind reviewing in economics journals and find that the gender gap in acceptance rates is lower when journals use double-blind reviewing. In an experiment of single-blind versus double-blind reviewing, Blank (1991) finds women fare slightly better under a double-blind reviewing system, but the estimated effects are not significant. Abrevaya and Hamermesh (2012) find no evidence of gender discrimination or altruism based on the gender pairing of reviewers and authors in the review process at a top field journal, though the journal uses a double-blind review process. (Of course, reviewers are often able to determine the identity and gender of authors if the paper is posted online.) An important limitation of this gender-pairing research design, however, is that it may fail to identify gender bias in the peer review process if women and men both discriminate against female authors. Card, DellaVigna, Funk, and Iriberry (2018) study referee decisions at four leading economics journals and similarly find no evidence of differential gender bias among reviewers or editors. However, they show that both male and female referees appear to hold female authors to a higher standard (as measured by citation counts), resulting in a substantial difference in the probability that female-authored papers receive a revise and resubmit. Similarly, Grossbard, Yilmazer, and Zhang (2018) show that papers in demographic economics journals with female authors receive more citations. Hengel (2017) adds a different dimension to the evidence that higher editorial standards are imposed on women in economics. She finds that economic research papers written by female authors spend six months longer under review at one top journal, although female-authored papers are more readable (using five different measures of writing clarity) and the gender gap in readability grows over the peer-review process. Hamermesh (2013) finds that, regardless of the reason, female authors have been substantially

underrepresented in top journals since the 1980s. While the evidence is not conclusive, differences in coauthorship networks and potential bias in the publishing process may both contribute to this gap.

External recognition through conference participation may also serve as a barrier to success for women. Women are underrepresented at high-profile conferences in economics compared to the overall share of female assistant professors, which is important if tenure committees use these presentations as a measure of prestige or external recognition of quality work (Chari and Goldsmith-Pinkham 2017).

Finally, the evaluation process for tenure and promotion may systematically disadvantage women. Evidence has been accumulating that implicit bias, which can lead to discrimination on the basis of unconscious attitudes and associations, is a problem in academia and can affect both hiring and promotion decisions on many margins (for a discussion in this journal, see Bayer and Rouse 2016). For example, faculty evaluating curriculum vitae with randomly assigned names are more likely to positively evaluate and hire male applicants for tenure-track jobs (Steinpreis, Anders, and Ritzke 1999). Letters of recommendation written for individuals applying for academic positions use different adjectives to describe men and women, and the characteristics used to describe women are viewed more negatively in hiring decisions (Madera, Hebl, and Martin 2009; Schmader, Whitehead, and Wysocki 2007). More specific to economics, equally productive female economists in Italy are less likely to be promoted to associate or full professor when randomly assigned to an all-male promotion committee, but there is no gender gap when women are assigned to a mixed-gender committee (De Paola and Scoppa 2015).

Even policies that have been supported on the grounds of gender equity may create biases against women's success. Antecol, Bedard, and Stearns (2018) examine the effect of gender-neutral tenure-clock stopping policies, which allow assistant professors who have children to extend their tenure clock. They find that such policies substantially increase the probability that men get tenure in their first job, but reduce the probability that women get tenure. Observed publishing outcomes suggest that men use the additional time on the tenure clock to continue to work and publish while women do not. Moreover, this study also finds that a large and significant gap in the probability of tenure remains even when controlling for the number of publications in top-five and non-top-five journals.

Evidence of gendered expectations of performance exists in many other high-skilled occupations as well. In a study of physician referral practices, Sarsons (2017b) finds that female surgeons are more heavily penalized for negative patient outcomes, while male surgeons are more strongly rewarded after positive outcomes. Another study, of misconduct by financial advisors, finds female financial advisors engage in less-costly types of misconduct on average, but are also significantly more likely relative to men to face harsh punishments following misconduct (Egan, Matvos, and Seru 2017). Finally, men serving on promotion committees across academic disciplines evaluate female candidates less favorably when there are women on the committee as well (Bagues, Sylos-Labini, and Zinovyeva 2017). The contrast between economics and other academic disciplines in the lack of progress

that has been made in reducing gender inequalities, however, suggests that biases within institutions of economics may be particularly pervasive.

Discussion

Following the considerable growth in women's representation among economics students and faculty during the 1970s and 1980s, progress has leveled off in the last two decades. Economics has made less headway than the science, technology, engineering, and mathematics fields in terms of increasing the share of female undergraduate majors and PhD recipients (Bayer and Rouse 2016), which will make it even more difficult to close the faculty gender gap in economics going forward. Furthermore, common explanations for female academic disadvantage, such as heavier domestic responsibilities and an aversion to math intensity, fail to explain why economics is falling behind these other fields in terms of female persistence and promotion probabilities. What can explain the unique challenges that women seem to face in economics?

An adversarial and aggressive culture within academic economics is often advanced as a causal force in women's stalled progress in the profession, though its impact is difficult to quantify. Economics seminars, for example, have a reputation for being particularly hostile environments. The culture of an academic discipline can have gendered implications if women either fail to fully adapt to the culture or if they receive differential treatment as a result of it. Female economists appear to be less likely to engage in practices that are positively correlated with professional success, suggesting an inability or unwillingness to adapt to professional norms. For example, male academics self-cite more than female academics in many fields, but the male-to-female self-cite ratio is twice as high and more persistent in economics (King, Bergstrom, Correll, Jacquet, and West 2017). Applied economics fields attract a higher proportion of women, but this work is still seen by some as less rigorous or less important than traditionally male-dominated topics. Anecdotal evidence suggests that women may choose to go into less-male-dominated fields or leave academia altogether based on early experiences with toxic environments that men are more likely to tolerate.

It is obviously difficult to obtain quantitative estimates of the extent of outright harassment of women in economics. We do know that there are many reports of women in economics experiencing inappropriate behavior in job interviews, seminars, meetings, and at conferences (Shinall 2018). In addition, the language used to describe female economists on at least one anonymous online forum is often sexual and derogatory, in a way that it is not for men (Wu 2017). Recent evidence suggests that gender harassment is a problem in academics more broadly (National Academies of Sciences, Engineering, and Medicine, 2018). Such behavior is often normalized and tolerated in male-dominated settings, making it difficult to change. Thus, the National Academies of Sciences offer several evidence-based recommendations to address harassment in the university setting that may be directly relevant

to economics. In particular, they advise reducing the importance of hierarchical relationships and implementing “power-diffusion” mechanisms such as mentoring networks. They also argue that taking explicit actions to achieve greater gender equity in the hiring and promotion process is an essential step in creating a diverse and respectful environment.

The evidence summarized above suggests two primary mechanisms through which the barriers against women in economics may operate: differences in productivity between men and women, and differences in how they are evaluated. Women may be on average less productive than men due to childbearing and other family responsibilities, a higher propensity to engage in service activities instead of research, or differences in the type of research in which they choose to invest their time. The distinct experiences of men and women in the profession may also contribute to productivity gaps that arise as a result of differences in collaborative networks, access to mentors, and gender harassment. But gender gaps conditional on productivity are also larger in economics than in other academic disciplines, suggesting that a second factor explaining female disadvantage in economics may be disparate assessment of men and women. It appears that women are held to higher standards than men of equal ability, and need to publish more, higher-quality work to achieve equal levels of success in this profession.

Continued progress toward equality in academic economics will require a widespread awareness that these barriers exist, accompanied by a concerted effort to remove opportunities for bias in the hiring and promotion process. However, first steps have been slow in coming. A 2008 survey of AEA members found, in addition to substantial differences in the policy views of male and female economists, a meaningful gender gap in their beliefs on equal opportunity in the profession (May, McGarvey, and Whaples 2014). While 76 percent of female AEA members believed that opportunities for economics faculty in the US favor men, fewer than 20 percent of men shared the same view. In fact, one-third of male economists felt that opportunities in economics actually favor women. To the extent that such beliefs persist, they are a major obstacle to the development of new diversity initiatives.

Diversifying the economics profession is important, because a greater breadth of individual perspectives will affect what is taught in the classroom, what research questions are asked, and how policy discussions are addressed. In addition, to the extent that women’s stalled progress in economics is the result of discrimination or biased assessment, as recent evidence suggests, continued action to remove these barriers can be justified both on the basis of simple fairness and also on the benefits of creating an environment where equal work yields equal rewards.

■ *We are grateful to Juliana Helo and Sangeetha Ramamurthy for excellent research assistance, and to Dick Startz and Meredith Startz for helpful comments.*

References

- Abrevaya, Jason, and Daniel S. Hamermesh.** 2012. "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?" *Review of Economics and Statistics* 94(1): 202–207.
- Antecol, Heather, Kelly Bedard, and Jenna Stearns.** 2018. "Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?" *American Economic Review* 108(9): 2420–41.
- Babcock, Linda, Maria P. Recalde, Lise Vesterlund, and Laurie Weingart.** 2017. "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability." *American Economic Review* 107(3): 714–47.
- Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva.** 2017. "Does the Gender Composition of Scientific Committees Matter?" *American Economic Review* 107(4): 1207–38.
- Bartlett, Robin L.** 1998. "CSWEP: 25 Years at a Time." *Journal of Economic Perspectives* 12(4): 177–183.
- Bayer, Amanda, and Cecilia Elena Rouse.** 2016. "Diversity in the Economics Profession: A New Attack on an Old Problem." *Journal of Economic Perspectives* 30(4): 221–42.
- Bell, Carolyn Shaw.** 1973. "Report of the Committee on the Status of Women in the Economics Profession." *American Economic Review* 63(2): 508–11.
- Beneito, Pilar, J. E. Boscá, Javier Ferri, and Manu García.** 2018. "Women across Subfields in Economics: Relative Performance and Beliefs." FEDEA Working Paper No 2018-06.
- Bergmann, Barbara R.** 1998. "Two Cheers for CSWEP?" *Journal of Economic Perspectives* 12(4): 185–189.
- Blank, Rebecca M.** 1991. "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from *The American Economic Review*." *American Economic Review* 81(5): 1041–67.
- Blau, Francine D., Janet M. Currie, Rachel T. A. Croson, and Donna K. Ginther.** 2010. "Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial." *American Economic Review* 100(2): 348–52.
- Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberry.** 2018. "Are Referees and Editors in Economics Gender Neutral?" October https://eml.berkeley.edu/~sdellavi/wp/EditorGender_Oct24.pdf.
- Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams.** 2014. "Women in Academic Science: A Changing Landscape." *Psychological Science in the Public Interest* 15(3): 75–141.
- Chari, Anusha, and Paul Goldsmith-Pinkham.** 2017. "Gender Representation in Economics across Topics and Time: Evidence from the NBER Summer Institute." NBER Working Paper 23953.
- Cherrier, Beatrice.** 2017a. "Classifying Economics: A History of the JEL Codes." *Journal of Economic Literature* 55(2): 545–79.
- Cherrier, Beatrice.** 2017b. "'The American Economic Association Declares that Economics is not a Man's Field': The Missing Story." *The Undercover Historian* (blog), March 31, 2017.
- Cherrier, Beatrice, Cleo Chassonnery-Zaigouche, and John Singleton.** 2018. "A Game of Mirrors? Economists' Models of the Labor Market and the 1970s Gender Reckoning." *The Undercover Historian* (blog), March 6, 2018.
- CSWEP.** 1973. "Combating Role Prejudice and Sex Discrimination: Findings of the American Economic Association Committee on the Status of Women in the Economics Profession." *American Economic Review* 63(5): 1049–61.
- CSWEP.** 2018. "2017 Report on the Status of Women in the Economics Profession." *CSWEP News*, Issue I. <https://www.aeaweb.org/content/file?id=6388>.
- De Paola, Maria, and Vincenzo Scoppa.** 2015. "Gender Discrimination and Evaluators' Gender: Evidence from Italian Academia." *Economica* 82(325): 162–88.
- Ductor, Lorenzo, Sanjeev Goyal, and Anja Prummer.** 2018. "Gender & Collaboration." Cambridge-INET Working Paper 1807.
- Egan, Mark L., Gregor Matvos, and Amit Seru.** 2017. "When Harry Fired Sally: The Double Standard in Punishing Misconduct." NBER Working Paper 23242.
- Ferber, Marianne A., and Michelle Teiman.** 1980. "Are Women Economists at a Disadvantage in Publishing Journal Articles?" *Eastern Economic Journal* 6(3/4): 189–193.
- Forget, Evelyn L.** 2011. "American Women and the Economics Profession in the Twentieth Century." *Æconomia* 1(1): 19–30.
- Fourcade, Marion, Etienne Ollion, and Yann Algan.** 2015. "The Superiority of Economists." *Journal of Economic Perspectives* 29(1): 89–114.
- Friedman, Milton.** 1998. "A Comment on CSWEP." *Journal of Economic Perspectives* 12(4): 197–99.
- Ginther, Donna K., and Shulamit Kahn.** 2004.

- "Women in Economics: Moving Up or Falling Off the Academic Career Ladder?" *Journal of Economic Perspectives* 18(3): 193–214.
- GINther, Donna K., and Shulamit Kahn.** 2009. "Does Science Promote Women? Evidence from Academia 1973–2001." Chap. 5. in *Science and Engineering Careers in the United States*, edited by Richard B. Freeman and Daniel F. Goroff. Chicago, IL: University of Chicago Press for NBER Science Engineering Workforce Project.
- GINther, Donna K., and Shulamit Kahn.** 2014. "Academic Women's Careers in the Social Sciences." Chap. 11 in *The Economics of Economists: Institutional Setting, Individual Incentives, and Future Prospects*, edited by Alessandro Lanteri and Jack Vromen. Cambridge University Press.
- Grossbard, Shoshana, Tansel Yilmazer, and Lingrui Zhang.** 2018. "The Gender Gap in Citations: Lessons from Demographic Economics Journals." HCEO Working Paper 2018-078.
- Hamermesh, Daniel S.** 2013. "Six Decades of Top Economics Publishing: Who and How?" *Journal of Economic Literature* 51(1): 162–72.
- Hengel, Erin.** 2017. "Publishing While Female. Are Women Held to Higher Standards? Evidence from Peer Review." Cambridge Working Papers in Economics 1753, Faculty of Economics, University of Cambridge.
- King, Molly M., Carl T. Bergstrom, Shelley J. Correll, Jennifer Jacquet, and Jevin D. West.** 2017. "Men Set Their Own Cites High: Gender and Self-Citation across Fields and Over Time." *Socius* 3.
- Madera, Juan M., Michelle R. Hebl, and Randi C. Martin.** 2009. "Gender and Letters of Recommendation for Academia: Agentic and Communal Differences." *Journal of Applied Psychology* 94(6): 1591–99.
- May, Ann Mari, Mary G. McGarvey, and Robert Whaples.** 2014. "Are Disagreements among Male and Female Economists Marginal at Best?: A Survey of AEA Members and Their Views on Economics and Economic Policy." *Contemporary Economic Policy* 32(1): 111–132.
- McDowell, John M., Larry D. Singell Jr, and Mark Stater.** 2006. "Two to Tango? Gender Differences in the Decisions to Publish and Coauthor." *Economic Inquiry* 44(1): 153–68.
- McDowell, John M., Larry D. Singell Jr, and James P. Ziliak.** 2001. "Gender and Promotion in the Economics Profession." *ILR Review* 54(2): 224–44.
- Misra, Joya, Jennifer Hickee Lundquist, Elissa Holmes, and Stephanie Agiomavritis.** 2011. "The Ivory Ceiling of Service Work." *Academe* 97(1): 22–26.
- National Academies of Sciences, Engineering, and Medicine.** 2018. *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*. Consensus Study Report. Washington, DC: The National Academies Press.
- Nelson, Donna J.** 2004. "Nelson Diversity Surveys." Norman, OK: Diversity in Science Association. <http://drdonnajnelson.oucreate.com/diversity/top50.html> (accessed June 3, 2018).
- Sarsons, Heather.** 2017a. "Gender Differences in Recognition for Group Work." https://scholar.harvard.edu/files/sarsons/files/full_v6.pdf.
- Sarsons, Heather.** 2017b. "Interpreting Signals in the Labor Market: Evidence from Medical Referrals." https://scholar.harvard.edu/files/sarsons/files/sarsons_jmp.pdf.
- Schmader, Toni, Jessica Whitehead, and Vicki H. Wysocki.** 2007. "A Linguistic Comparison of Letters of Recommendation for Male and Female Chemistry and Biochemistry Job Applicants." *Sex Roles* 57(7–8): 509–514.
- Shinall, Jennifer Bennett.** 2018. "Dealing with Sexual Harassment." *CSWEP News*, Issue I. <https://www.aeaweb.org/content/file?id=6769>.
- Steinpreis, Rhea E., Katie A. Anders, and Dawn Ritzke.** 1999. "The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study." *Sex Roles* 41(7–8): 509–528.
- Stevenson, Betsey, and Hanna Zlotnick.** 2018. "Representations of Men and Women in Introductory Economics Textbooks." *AEA Papers and Proceedings* 108(5): 180–85.
- Strober, Myra.** 2016. *Sharing the Work: What My Family and Career Taught Me about Breaking Through (and Holding the Door Open for Others)*. MIT Press.
- Wu, Alice H.** 2017. "Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum." Working Papers 2017-09, Princeton University, Woodrow Wilson School of Public and International Affairs, Center for Health and Wellbeing. <https://drive.google.com/file/d/0BwjFN4HbBrDBbnFqZzdLWThDb0U/view>.
- Xie, Yu, and Kimberlee A. Shauman.** 2003. *Women in Science: Career Processes and Outcomes*. Cambridge: Harvard University Press.

Variation in Women’s Success across PhD Programs in Economics

Leah Boustan and Andrew Langan

In 2017, women made up 32 percent of entering PhD students in economics. The share of women in economics is below many other fields including science, technology, engineering, and mathematics and has not increased since the 1990s (as discussed in this journal by Bayer and Rouse 2016). This paper adds new data—both quantitative and qualitative—on graduate programs in economics to understand the wide and persistent variation in women’s success across departments. We then use these insights to suggest “best practices” for department chairs, PhD admissions committees, and others hoping to increase the number of women in economics.

Our quantitative data come from two sources: newly available annual surveys of graduate departments conducted by the Committee on the Status of Women in the Economics Profession (CSWEP), and our own hand-collected faculty rosters from PhD-granting economics departments in the United States from 1994 to 2017. The CSWEP survey contains information on the number of men and women graduating from each program by year, and the number of job market placements by job type and gender.¹ We organize our roster data to associate each faculty member with their graduation cohort and alma mater to examine other early career outcomes, such as placement rank, publications, and promotion.

¹ CSWEP’s data are now posted on ICPSR (Study Number 37118); we received an early copy of the data for this paper. We thank Shelley Lundberg and Margaret Levenstein for access to the data. By agreement with CSWEP, names of the institutions are suppressed.

■ *Leah Boustan is a Professor of Economics and Andrew Langan is a PhD student, both at Princeton University, Princeton, New Jersey. Boustan is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Boustan is the corresponding author at lboustan@princeton.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at

The opening four sections of this paper document a number of facts. First, there was wide variation in the average share of women in the graduating classes of economics PhD programs during the past 30 years, ranging from 10 percent to more than 50 percent of the class. This variation is primarily explained by differences in initial admission, rather than differential attrition from the program. Many programs, particularly the largest ones, increased the share of women on the student body from the 1990s to the 2010s. Yet the share of women in a PhD program tends to be a persistent attribute of a department, with a strong correlation in the gender composition of graduating classes over time.

Second, we show that departments with a greater share of women on their faculty also have more women in their student body: a 10 percentage point increase in faculty share is associated with a 2.5 percentage point increase in student share. This relationship could be causal if, for example, women on the faculty serve as role models for women students, or it could reflect other departmental attributes that are attractive to both women faculty members and graduate students.

Third, we document that, on average, men and women who graduated from the same program between 1994 and 2017 are no different in their propensity to be offered and accept a faculty position at a US PhD-granting department or to be promoted to associate professor within ten years of graduation. But conditional on taking a job in a US PhD-granting economics department, men land placements at higher-ranked departments and publish more in the top journals in the first seven years after obtaining their degree.

Fourth, we rank 22 large (anonymized) programs on the gender gap in these early career outcomes. We focus on programs with sufficient data for both men and women; these 22 programs train two-thirds of the faculty at PhD programs in the United States. Our ranking identifies large variation in relative success for women across graduate programs in outcomes like job placement, publication, and promotion. For example, women graduating from departments with better relative outcomes have 9 percent higher placement rates than men, while women graduating from departments with worse relative outcomes have 8 percent lower placement rates than men. Yet in all cases, men are more likely than women to publish in top journals, suggesting that women face a common set of impediments in their early careers regardless of their graduate institution.

This ranking across departments guided our selection of departments for a set of structured qualitative interviews designed to learn more about variation in the mechanics and culture of graduate instruction in economics. We conducted 31 interviews with faculty members and former students at five programs—two that achieved good relative outcomes for women, two that achieved poor relative outcomes for women, and one in the middle of the pack. All interviews were conducted and transcribed by Leyla Mocan, a Master in Public Policy student at the Woodrow Wilson School at Princeton. Our interviews confirm that having women on the faculty inspires women in the student body to succeed. The interviews also uncover several features of graduate programs that are associated with good relative outcomes for women: structuring the graduate program to formalize

key aspects of advisor–student contact; creating a collegial atmosphere in research seminars; and developing awareness of gender issues, especially by senior male faculty. We see many of these ideas as “gender neutral,” in the sense that they would likely improve the climate of graduate instruction for all students regardless of gender. However, our interviews suggest that these reforms would likely have a disparate impact on women.

The Share of Women Graduating from Economics PhD Programs

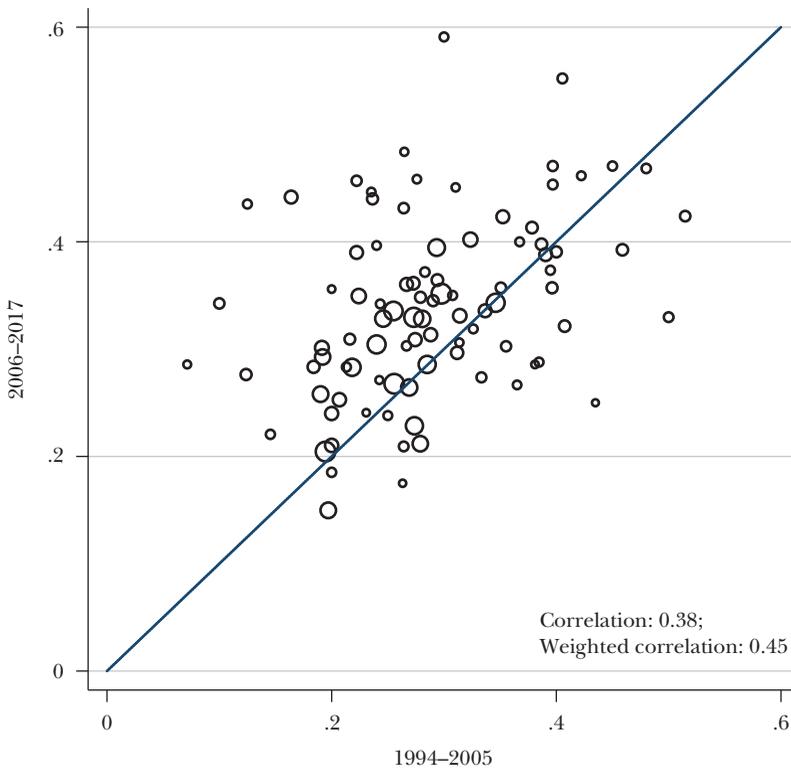
This section presents new facts about variation in the share of women in graduating classes from economics PhD programs. We limit our analysis to the 88 out of 127 economics departments with a large enough number of entering and graduating PhD students who responded to the CSWEP survey.² We aggregate the annual data into two broader graduation cohorts: students graduating between 1994 and 2005 and between 2006 and 2017. These periods divide our data in half.

Figure 1 illustrates that the share of women in the graduate student body ranges from 10 percent to more than 50 percent. The share of women in a program is a persistent attribute, with a correlation between the two time periods of 0.38 (0.45 if weighting by size of the program, and 0.53 when focusing on programs above the median size). Figure 2 shows a strong relationship between the share of women in entering first-year cohorts (1994–2012) and in graduating cohorts (1999–2017). The coefficient from regressing share of women in the graduating class on share of women in the entering class cannot be statistically distinguished from one, suggesting that variation in share of women across programs is not driven by differential attrition, such as gender differences in the probability of leaving the program after failing general exams. The year groupings in this figure are selected to match entering first-year cohorts at least approximately with corresponding graduating classes. We assume that students take around five years to graduate, but the result is robust to alternative assumptions.

Correlations with Observable Departmental Attributes

Differences in the share of women graduating from PhD departments in economics are correlated with some observable characteristics of these programs. Table 1 considers the association between share of women in the graduating class and six attributes of a PhD program: its *US News & World Report* department ranking; the share of women on its faculty; the field mix of its faculty; the average

²We focus on departments that were above the 25th percentile in the number of entering first-year students (123) and the number of graduates (68) reported to CSWEP during the full period, and for which we have faculty rosters, to permit other analyses.

*Figure 1***Share Women PhD Recipients by Department: 1994–2005 versus 2006–2017**

Source: CSWEP department survey.

Note: This figure plots women's share of economics PhD recipients from 2006 to 2017 against their share from 1994 to 2005, by department. Points are weighted by the total number of graduates over the entire period per department. We exclude years with imputed counts of PhD graduates and large outliers from department totals.

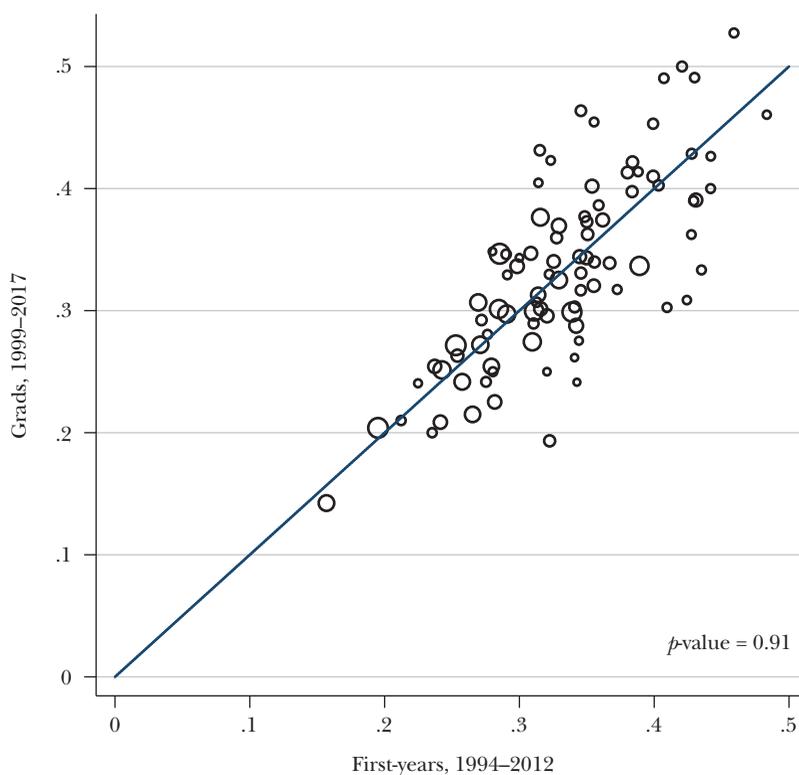
size of its entering class; the share of non-US-citizen entering students, and the institution's location in a small, medium, or large city.

The first six columns enter each attribute separately and the last column includes all control variables that were individually statistically significant. With 88 observations, we have limited power to detect relationships. Yet we find that the share of women in the graduating class rises in a clear way with the share of women on the faculty. The final column suggests that a 10 percentage point increase in the share of women on the faculty (roughly equivalent to the interquartile range) is associated with a 2.5 percentage point increase in the share of women in the graduating class. This positive association is consistent with Hale and Regev (2014), which uses retirements of men to consider exogenous changes in share of women on the faculty.

A few other departmental characteristics influence the gender composition of the student body. Programs located in the country's 15 largest metro areas have

Figure 2

Share Women, PhD Recipients versus First-Years, by Department, Entering Classes of 1994–2012



Source: CSWEP department survey.

Note: This figure plots women's share of economics PhD recipients from 1999 to 2017 against their share of first-year students from 1994 to 2012, by department. The p -value shown is taken from a t -test of the null hypothesis that the share of women is equal in the two groups within department. Points are weighted by the total number of graduates over the entire period per department. We exclude years with imputed counts of PhD graduates or first-years and large outliers from department totals.

more women in the student body, as do departments further down the 2017 *US News* ranking. Faculty research field mix, program size, and the share of non-US students, as measured by the National Science Foundation's *Survey of Graduate Students and Postdoctorates in Science and Engineering*, do not seem to affect gender balance. In other calculations not shown here, we also see no association between any of these department attributes and the relative placement rates of women.³

To measure gender and field composition of faculty, we compile data on academic rosters across economics departments from three sources: the *Prentice*

³Neumark and Gardecki (1998) and Hilmer and Hilmer (2007) likewise find no positive association in the late 1970s through early 1990s between the share of women on the faculty and students' early career outcomes.

Table 1

Effect of Department Characteristics on Share Women in Graduating Class, Economics PhDs

| Department characteristics | Mean | Share women graduates | | | | | | |
|-----------------------------|--------|-----------------------|---------------------|-------------------|---------------------|-------------------|-------------------|--------------------|
| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Rank 11–20 | 0.102 | 0.024 (0.034) | | | | | | 0.035 (0.034) |
| Rank 21–30 | 0.159 | 0.049 (0.031) | | | | | | 0.060* (0.032) |
| Rank 31+ | 0.625 | 0.080*** (0.025) | | | | | | 0.079** (0.036) |
| Faculty Share Women | 0.156 | | 0.409*** (0.127) | | | | | 0.245* (0.130) |
| Faculty Share Applied Micro | 0.389 | | | 0.231 (0.257) | | | | |
| Faculty Share Theory | 0.228 | | | -0.090 (0.254) | | | | |
| Faculty Share Macro | 0.252 | | | 0.007 (0.258) | | | | |
| Average Cohort Size | 15.961 | | | | -0.003** (0.001) | | | 0.000 (0.002) |
| Share Foreign Entrant | 0.472 | | | | | -0.006 (0.061) | | |
| Medium City | 0.375 | | | | | | 0.032 (0.020) | 0.023 (0.019) |
| Large City | 0.330 | | | | | | 0.037* (0.021) | 0.050** (0.021) |
| Constant | | 0.272*** 0.023 | 0.268*** 0.021 | 0.261 0.191 | 0.383*** 0.021 | 0.335*** 0.030 | 0.308*** 0.015 | 0.214** 0.056 |
| Observations | | 88 | 88 | 88 | 88 | 88 | 88 | 88 |
| R ² | | 0.136 | 0.108 | 0.027 | 0.073 | 0.000 | 0.042 | 0.243 |

Source: The sample includes 88 departments that responded to the CSWEP survey, exceeded the 25th percentile of both the number of entering and graduating students among respondents, and for which we have departmental rosters available from Langan (2018). Economics department rank comes from the 2017 edition of *US News and World Report*. Faculty share women and average cohort size come from 1994–2016 CSWEP responses. Field composition is obtained from our academic roster (Langan 2018) by counting total person-years in the department by individuals' stated fields, using JEL codes from the 1997 AEA members survey, the *Prentice Hall Guide to Economics Faculty*, 1994–2006, and keyword matching from statements on individual and department websites. Share of foreign first-year students comes from 1994–2017 department responses to the National Science Foundation's *Survey of Graduate Students and Postdoctorates in Science and Engineering*.

Note: This table reports the results of regressing women's share of PhD recipients from economics departments on various department characteristics: its *US News & World Report* department ranking; the share of women on its faculty; the field mix of its faculty; the average size of its entering class; the share of non-US-citizen entering students; and the institution's location in a small, medium, or large city. Column 6 regresses share women on the size of the city in which the department is located: the top 15 metro areas are considered "large" and smaller areas with more than 150,000 population are considered "medium." When calculating women's share of graduates from each department in the CSWEP survey data, we drop department-year observations with imputed values for graduates or academic placements at PhD granting institutions and large outlier values.

***, **, and * represents statistical significance at the 1, 5, and 10 percent levels, respectively.

Hall Guide to Economics Faculty, published in seven editions between 1994 and 2006; departmental and faculty web pages accessed using the Wayback Machine (www.archive.org); and the 1989 and 1993 biographical surveys of members by the American Economic Association. More details on construction of the faculty roster are available in Langan (2018).

Early Career Outcomes by Gender

We now turn to examining differences in early career outcomes of recent economics PhDs by gender. We calculate a set of relative ratios for men and women graduating from the same programs for outcomes like academic placement, placement rank, and so on. We then ask in the next section how those ratios vary across departments. Here, we start by identifying two average tendencies: 1) women and men who graduated from the same department are equally likely to place into US doctoral departments; and 2) conditional on receiving an academic placement, women place into lower-ranked departments and publish fewer papers in the first seven years of their career.

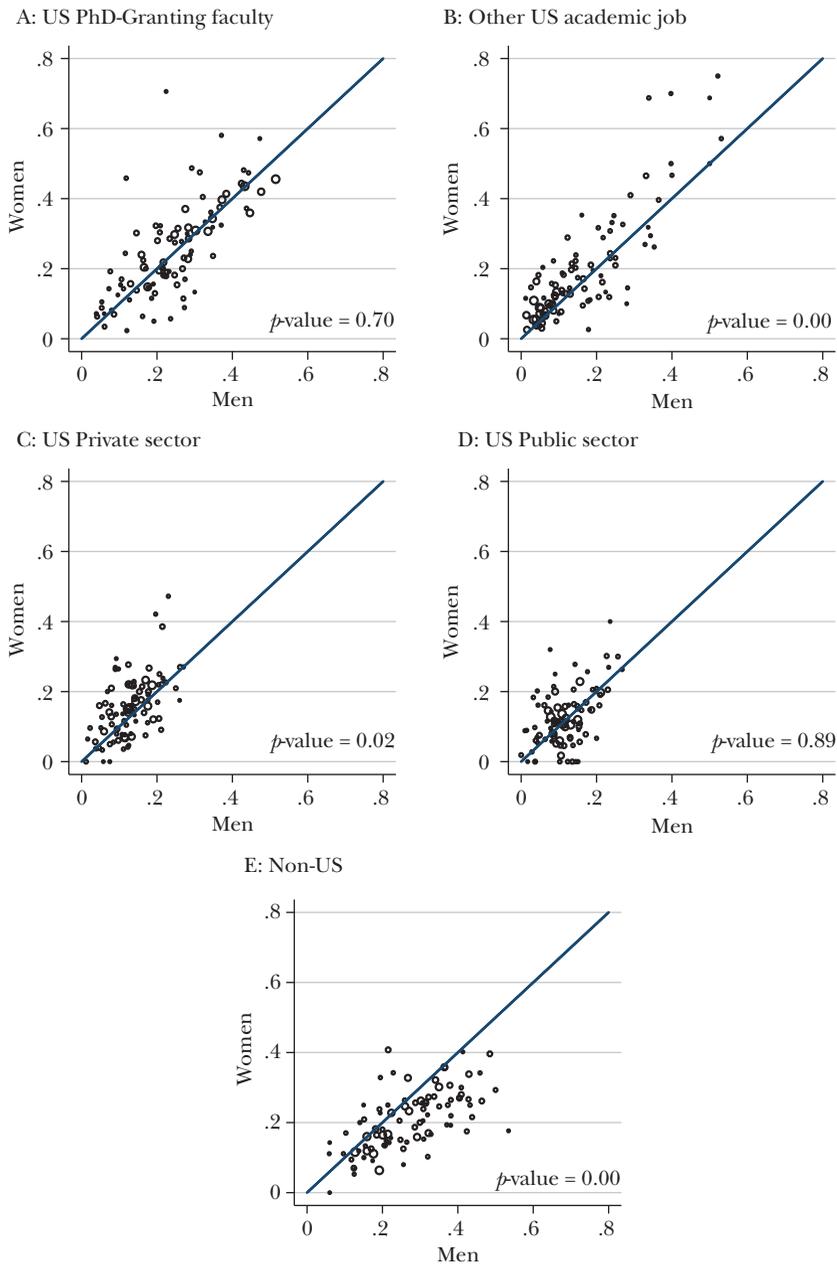
Figure 3 graphs the job placement rate by gender for graduates of the 88 PhD programs in our core sample from 1994 to 2017. We consider five placement categories: faculty jobs in US PhD-granting departments, other US academic jobs, US private sector jobs, US public sector jobs, and jobs outside of the United States. The p -values under each panel refer to the null hypothesis that rates of placement are equal for men and women (a small p -value indicates strong evidence against the null hypothesis). We find *no* statistically significant difference in the average placement rate of men and women graduates from the same department into faculty jobs in US PhD-granting institutions, as shown in the first panel. In contrast, women are more likely to accept other US academic jobs, and men are more likely to take jobs outside of the United States.

To develop more detailed early career outcome measures, we move from CSWEP's job placement data to our faculty rosters. We create ratios of men and women who graduated from each department by associating each faculty member in our roster with their department of graduate instruction. For example, we observe one of our coauthors, Leah Boustan, in her first job placement in the UCLA Department of Economics in 2006, but we associate her outcomes (including her placement at UCLA) to her graduation cohort at Harvard University. We use the reshaped roster to define additional outcomes for 22 departments with at least ten women placed into PhD-granting economics departments in order to calculate reasonable averages. Because our rosters include only academic economics departments and not, for instance, business, public policy, or resource economics programs, they constitute only a subset of the CSWEP placements, but the two placement rates are highly correlated overall, and by gender (correlation = 0.75).

Figure 4 compares a set of early career outcomes for men and women who graduated from each of the 22 departments. Men place at departments ranked higher

Figure 3

Job Placement Rates by Gender and Department, Five Sectors

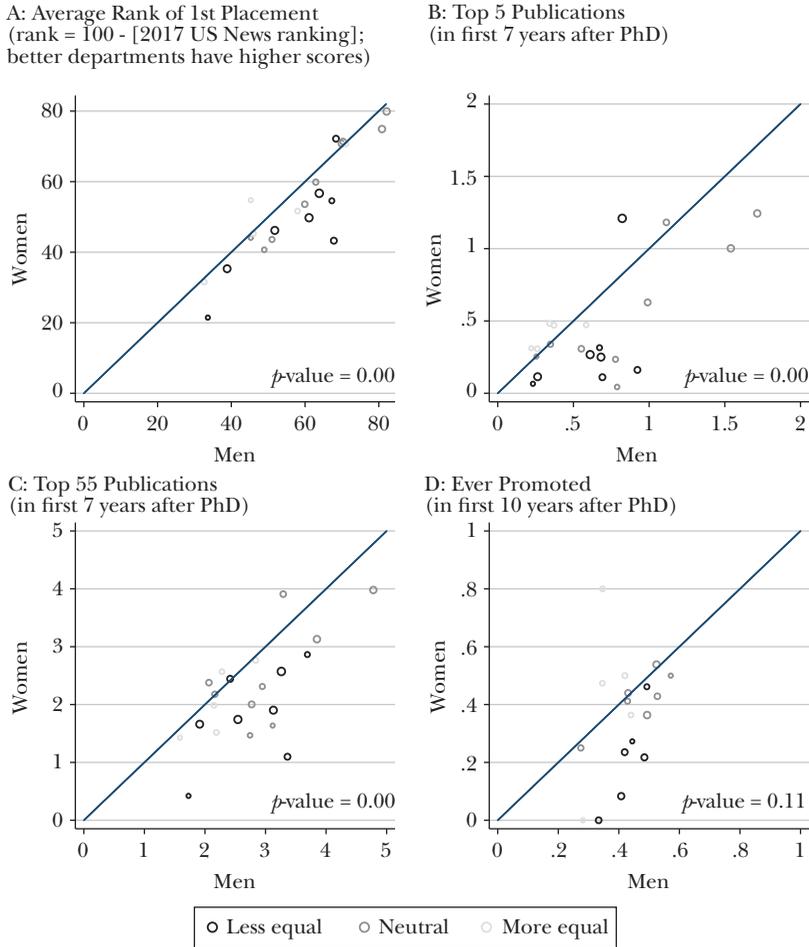


Source: CSWEP department survey.

Note: This figure plots women's versus men's rate of placement into various types of job, for economics PhD recipients in 1994–2017, by department (indicated by circles). The p -values shown refer to a test of the null hypothesis that placement rates for a given job type are equal for men and women (a small p -value indicates strong evidence against the null hypothesis.) “US PhD-Granting Faculty” refers to tenure-track faculty jobs in a department that awards doctoral degrees. Points are weighted by the total number of graduates over the entire period per department. We exclude years with imputed counts of PhD graduates or academic placements at PhD granting institutions and large outliers from department totals.

Figure 4

Post-Graduation Outcomes for Men and Women, PhD Economists by Graduate Department



Source: Source: Faculty rosters, Langan (2018).

Note: This Figure plots, for various graduate departments, women's average outcome against men's for four early-career outcomes. We focus on PhD graduates from 22 large US economics departments for which we can identify at least 10 women placed from the 1987 through 2017 graduating cohorts into US economics departments with doctoral programs (or for Figure 4D, 18 graduate departments with 10 women placed between 1987 and 2010). In Figure 4A, we plot the average rank for the first economics department where individuals place after graduation. (Departments not ranked by US News are assigned a *US News* rank of 100). In Figure 4B, for each graduate department, we plot the average number of publications for men versus women graduates in a Top 5 journal (*American Economic Review* (excluding *Papers and Proceedings*), *Quarterly Journal of Economics*, *Econometrica*, *Journal of Political Economy*, and *Review of Economic Studies*) in the first 7 years after PhD. In Figure 4C, for each graduate department, we plot number of publications for men and women in journals ranked 55 or better in the RePEc aggregate journal rankings in the first 7 years after a PhD. For post-2010 graduates, both top 5 and top 55 total publications are predicted for those with at least one publication. In Figure 4D, for each graduate department, we plot the share of men and women, graduating ever observed as an associate or full professor in our faculty rosters within 10 years after receiving their PhD. The p -values are taken from a t -test of the null hypothesis that men and women's average outcomes are equal within department (a small p -value indicates strong evidence against the null hypothesis). Points are weighted by the total number of graduates over the entire period per department. Dots shaded lighter indicate graduate departments that are more equal in terms of having better relative outcomes for women, as summarized in Table 2.

by *US News* and have more publications within their first seven years of graduation. The average gap in placement rank between men and women graduates of the same program is 4.8 rank points, as shown in Figure 4A. Figures 4B and 4C show that men publish more papers in Top 5 or Top 55 economics journals within seven years of graduation; publications are culled from the table of contents data provided by journals to *RePEc*. The average gap in Top 55 publications between men and women graduates of the same program is 0.58.⁴ Figure 4D shows that women are marginally less likely to be promoted to associate or full professor at a PhD-granting economics department within ten years of graduation, relative to men from their alma mater.⁵ Balanced promotion rates by gender could arise if men publish more than women but also face a higher publication bar for tenure, given their higher-ranked first placement.

Documenting Differences in Relative Outcomes for Women across Departments

We are particularly interested in documenting the variation across departments in relative outcomes for women, including graduation rates, placement into an academic job, and later research performance. In Table 2, we rank 22 anonymized graduate programs from “A,” the department with the best outcomes for women graduate students, relative to men in their programs, to “V,” the department with the worst relative outcomes. The ranking is derived by converting each of the outcomes into a z-score and then calculating the mean z-score across outcomes for each department; details are presented in Appendix A.

Average outcomes for three ranked categories (better & best; neutral; worse & worst) are reported in the last three rows of the table. Across these three categories, women make up a similar share of the student body (28 to 30 percent) and first-year students are equally likely to graduate from the program (column labeled “Retention”) regardless of gender. Differences arise at the job placement and assistant professor stages. At programs that are relatively better for women, women have 9 percent higher academic placement rates than men, placement ranks that are nearly identical to men in their program, and marginally lower numbers of top 55 publications. At programs that are relatively worse for women, women have 8 percent lower academic placement rates than men, place into substantially lower-ranked departments, and publish approximately 30 percent fewer papers in a top 55 journal.

⁴The relative publication records of men and women in our sample is consistent with Antecol, Bedard, and Stearns (2018) and Sarsons (2017) for economics, and with broader results across fields summarized in Ceci, Ginther, Kahn, and Williams (2014). For additional work on gender and publication in economics, see Blank (1991), McDowell, Singell, and Stater (2006), Abrevaya and Hamermesh (2012), Bransch and Kvasnicka (2017), and Hengel (2017).

⁵Ginther and Kahn (2004, 2015) find that men are more likely than women to receive tenure in economics, controlling for year of graduation and institutional quality of alma mater.

Table 2

Comparative Success of Women by Graduate Department

(icons represent a department's position in the distribution of each statistic: ✓✓ = top 10%; ✓ = top 25%; × = bottom 25%; ×× = bottom 10%).

| Group | ID | Share women | Change in share women | Retention | Rates of placement at PhD | Placement rank | Top 55 | Top 5 | Promotion |
|----------------------------------|-------|-------------|-----------------------|-----------|---------------------------|----------------|--------|-------|-----------|
| Best relative outcomes for women | A | ✓ | | × | ✓✓ | | ✓ | ✓✓ | |
| | B | | ✓ | ✓ | × | | | ✓ | ✓✓ |
| | C | | ✓✓ | | | ✓✓ | | ✓ | × |
| Better relative outcomes | D | | | | ✓ | | | ✓ | ✓ |
| | E | | ✓ | ✓✓ | | ✓ | | | ✓ |
| Neutral | F | | × | | | | ✓ | | |
| | G | | | ✓✓ | | | | | |
| | H | ✓ | | | | | × | | |
| | I | | | | | ✓ | | | |
| | J | | | | ×× | ✓ | | ✓✓ | |
| | K | × | × | | | × | ✓ | ✓✓ | |
| | L | | | | ✓ | ✓ | × | × | ×× |
| M | × | | | | | | | ✓ | |
| N | | | ✓ | | ✓✓ | | | × | |
| Worse relative outcomes | O | ×× | | | | | | ✓✓ | × |
| | P | | | | | | × | | |
| | Q | ✓✓ | ✓✓ | ✓ | ×× | ×× | ×× | × | |
| | R | ✓ | | × | | × | | | × |
| S | ✓✓ | × | × | × | | | | × | |
| Worst relative outcomes | T | ×× | ×× | | | ×× | ✓ | ×× | |
| | U | × | | | ×× | × | | | |
| | V | | ×× | ×× | | ✓✓ | ×× | × | |
| Better & Best | (A-E) | 0.30 | 1.28 | 1.02 | 1.09 | 1.01 | 0.92 | 1.21 | 1.14 |
| Neutral | (F-N) | 0.28 | 1.13 | 1.01 | 1.08 | 0.94 | 0.84 | 0.66 | 0.91 |
| Worse & Worst | (O-V) | 0.28 | 1.14 | 0.99 | 0.92 | 0.83 | 0.66 | 0.47 | 0.46 |

Note: This table ranks 22 graduate departments for which we can identify at least 10 women graduates who placed into an economics department with a doctoral program from 1987 to 2017. The ranking is conducted according to relative student outcomes for women versus men. Details for the ranking procedure are in Appendix A. Icons represent a department's position in the distribution of each statistic (✓✓ = top 10%; ✓ = top 25%; × = bottom 25%; ×× = bottom 10%). Averages for the statistics by group are shown in the lowest three rows. Share women is the fraction of first-year women over all first-year students for the whole period. For other statistics, graduate departments are ranked on relative rates, where a value greater than 1 reflects women outperforming men (or positive growth in the case of change in share women). Change in share women reports share women graduates (women graduates divided by total graduates) from 2006–2017 divided by share women graduates from 1993–2005 (see Figure 1). Retention is the share women graduates from 1999–2017 divided by share women first-year students from 1994–2012. This statistic compares the gender ratio at entrance and graduation, allowing 5 years for matriculation. Rates of placement at PhD is from CSWEP and is defined as for Figure 3. Placement rank, Top 55, Top 5, and Promotion are defined as for Figure 4.

Checkmarks in Table 2 indicate the departments that were in the top 10 percent (two checkmarks) or top 25 percent (one checkmark) of any given outcome (out of 22 departments), and ×'s indicate the departments that were in the bottom 10 percent (two ×'s) or bottom 25 percent (one ×). Departments that have better

relative outcomes for women on average are not *necessarily* better across the board. Even the best departments by our rankings have some poor relative outcomes for women (for example, Department C has low promotion rates), and even the worst departments have some good relative outcomes for women (Department T places women at highly-ranked departments relative to men).

Interviews about Why Some Departments Have Better Relative Outcomes for Women

Why are relative outcomes for women graduate students better at some departments than at others? One natural approach to answering this question for us, as empirical economists, might have been to gather systematic data about each department (perhaps by conducting a survey) and then performing a quantitative analysis to determine which attributes of a program predict success for women. However, we only have complete outcome variables for 22 programs, and we do not have a clear *a priori* sense of which explanatory variables may matter, nor how to measure them.

Instead, we conducted a series of exploratory interviews. We selected six departments to include in our interview sample: two departments with better than average outcomes (Departments B and C), two departments with worse than average outcomes for women (Departments O and T), and two departments in the middle of the group (Departments F and K). For each of these departments, we started by contacting the current department chair for a short interview. From there, we developed a snowball sample, asking the chair for a few names of faculty members and former students to reach out to next. In the end, we spoke with six or seven graduates or faculty from five departments, primarily in June and July 2018. Due to idiosyncratic scheduling difficulties we dropped one department (Department F) from the sample. Otherwise, we had a 91 percent response rate for all interview subjects contacted. Our interview subjects were 55 percent women and 45 percent men. Interviews ranged from 8 to 45 minutes, with the typical conversation lasting 15 minutes.

Each interview consisted of nine questions. We asked respondents to describe student contact with advisors, seminar culture and job market preparation, and also included open-ended questions about women in economics; we report the standard interview script in Appendix B. All interview subjects were told that their answers would be confidential. In our discussion of qualitative patterns, we suppress all names of individuals and of institutions, and we redact a few details to prevent department or subject identification.

From the interview responses, we observe a series of differences between departments with better and worse relative outcomes for women graduate students that are worth keeping in mind for departments interested in improving gender diversity or balance in outcomes for their members. The departments with the best relative outcomes tended to have a commitment to hiring women onto the faculty; regular and transparent processes for student–advisor contact; a more collegial seminar culture; and a stronger general awareness of gender issues among senior faculty.

Many of these departmental policies would be helpful for all students, but our interviews suggest that they may also help to narrow gender gaps. In the remainder of this section, we have organized a series of representative quotations from our interviews, emphasizing points of difference along these dimensions between departments with better and worse relative outcomes for women graduate students.

Women on the Faculty

One obvious difference between departments with better and worse relative outcomes for women graduate students is their commitment to and success in hiring women on the faculty. Students trained at Departments B and C talked about how valuable it was to have been taught by women faculty, and to have informal interaction with women faculty outside of the classroom. A former student from Department C said: “Every semester we always had one female teaching us in the core first year classes. Starting with that introduction to the graduate department made me feel like ‘I can do this and be a woman.’”⁶ A faculty member at Department C agreed, touting the value of the “women in economics group, female faculty who have lunches and dinners and coffees with female graduate students, a group that has been around as long as I can remember.” One faculty member from Department B described how the program was able to build up a group of women faculty, emphasizing that “our dean puts a lot of emphasis on diversity. It could even amount to an extra position if we come up with an additional excellent female candidate.”

In contrast, interviewees from departments with worse outcomes for women graduate students noted—and often lamented—their historical lack of representation of women on their faculties. These departments were only recently beginning to prioritize hiring women or had limited plans to do so. At our median Department K, the faculty were concerned about how their historical lack of women faculty affected their graduate students. One faculty member said, “[W]ell it starts with hiring. I’ve seen over the years that role models are incredibly important. I know that in our [redacted] field, for example, we currently have five faculty members who are all male. And I think we are all nice guys and inviting and encouraging and try to be gender neutral, yet whenever we have a female [visitor or speaker], women graduate students flock to her much more readily, there’s a comfort level and a role model that’s very important.” A faculty member at Department T concurs, relaying: “I have a student who is a woman who graduated maybe ten years ago ... she thinks her experience would have been better if there were more senior female faculty, that she would have felt more comfortable. ... So we’re working on that, we make offers, but like everyone else in the senior market it’s really tough. ... [I]t’s hard to get acceptances.” Unlike Departments B and C (and even Department K), Department T has not made an intentional effort to hire women. Instead, the chair simply tries to ensure that women are not overlooked on the job market, saying: “I look at

⁶Throughout the paper, we use the terms “men and women” instead of “male and female” when discussing gender in order to emphasize that gender identity is a social phenomenon. However, we leave the quotations from our interview subjects as stated.

the women we didn't fly out and I want to make sure that they're not better than the worst men that we fly out. ... [We] pay attention to see if we're gender neutral, at least in our junior hiring."

Hiring women on the faculty strikes us as a concrete and low-cost approach to creating a productive learning environment that might encourage women to enroll in the graduate program. Our interview subjects also articulated clear mechanisms whereby women on the faculty might benefit graduate students by serving as role models and mentors.

Advisor Contact

In many other disciplines in the sciences and the humanities, prospective PhD students apply to work with a particular faculty member before admission. In economics, it is typical for prospective students to apply to the department as a whole without specifying an intended field. All of the programs in our interview set have a decentralized process for second- or third-year students to select dissertation advisors. Many students reported approaching faculty with whom they took a course in the first-year core or in second-year field courses. This student's experience at Department O is typical of the haphazard search process: "I got [Faculty Member X as my advisor] just by talking a lot in class and by doing reasonably well ... that distinguished me and made me seen as a relatively attractive person to work with."

Students and faculty speculated that the current *laissez faire* approach can be intimidating for many students. It may be particularly hard for students who are in the minority in economics departments (including women) who may fear the actual or perceived skepticism of faculty. One student at Department C reported: "I think there was some implicit bias. I remember in the first year of the PhD program, I got [a high score on a core exam]... and my husband did not do particularly well in the course. But I remember the professor who was teaching the course reached out to my husband and tried to encourage him to go into [the field] and didn't say anything to me."

One feature that distinguished departments with better relative outcomes for women was the mandatory and regular nature of student works-in-progress seminars—that is, public venues for fourth- and fifth-year students to present new research and gather comments from a group of committed faculty. According to a student from Department C, "you would actually enroll in a course that was your lunchtime workshop; you were required to attend and present at least once a quarter. The faculty were really good about going to those things. They came regularly, and they would give you comments."

In contrast, at the departments with worse relative outcomes for women in our sample, works-in-progress seminars are offered for student research presentations, but students are not required to attend. As one faculty member describes the process, "[Department T] doesn't have ... a centralized policy around making sure everyone presented once or twice a year. ... There are certainly students who aren't going to any lunch, or if they are going they are not presenting. We don't do record keeping." Another faculty member agrees, saying that "one or two students get almost

completely lost, they don't make appointments with advisors, and their advisors don't reach out. ... There are some students who are not into getting advice very much. Unless they're a genius and come up with some fantastic thing, they tend to get lost."

Offering regular public venues for student feedback may be especially important for women, who mentioned off-campus advising settings in which they felt uncomfortable or to which they were not invited. This concern arose at departments of all levels on our ranking. One student from Department B described how "one senior male faculty was known for having these weekly 'salons.' He would take a bunch of students out to a bar. A bunch of students would go, it routinely felt like an old boys' club. ... I know one woman who went, but I wouldn't have felt comfortable." A student from Department C reported that she missed advising opportunities because of her gender: "[My advisor] would often take graduate students for beers, and he would take the men one on one. When I was first working with him, he said 'we should just meet in the office, we shouldn't get a beer, it could be misconstrued or you might feel I'm taking advantage of you.' So, either my husband would have to tag along, which was really boring for him, or another male graduate student would have to come, but I wouldn't get as much one-on-one attention."

Given the highly decentralized nature of the advising process in economics, we think that all departments have some room to experiment with developing more formalized points of student-advisor contact. One of our interview departments recently established a procedure to match incoming first-year students with a faculty mentor to address the unique challenges of the first-year program. Another department is experimenting with third-year research advising groups, wherein students are paired with field-specific faculty members. The chair of this department acknowledges that "before, it was all decentralized, students just had to reach out" and expressed hope for the new structure. Other departments have instituted regular faculty meetings to assess student progress and ensure that no students are falling between the cracks. One model for how to regularize the various steps of graduate training is the recent and widespread efforts to formalize the process of preparation for the job market—including holding information sessions, conducting mock interviews, designating a faculty member to be placement coordinator, and so on.

Seminar Culture

Another theme that emerged from our interviews is that departments with better relative outcomes for women are reported to have a less aggressive and more constructive climate in their research seminars. A faculty member at Department B portrayed seminars as "not enormously aggressive ... we don't take someone apart for the sake of taking them apart." A student from Department C had a similar impression. "Like most economics departments, [the seminar] was aggressive, but it wasn't as aggressive as some I've been to. You would get through your first few slides without being asked 'why did you choose this title? This is stupid.' Especially with the graduate students. They would ask questions more gently like 'don't you think that this would be a problem with identification?' Compared to 'this is not how to identify it.' But they wouldn't sit and let you make mistakes."

Two of the three departments with worse outcomes for women were noted for having aggressive seminar styles. A faculty member from Department K said that “the word that comes to mind is combative, perhaps aggressive. People start to talk about their work and much of their audience seems to think it’s their job to find the faults and tear it all down.” A student trained at Department O recalled that seminars were “fiercely competitive.” On the other hand, Department T was described as having seminars that are “pretty polite.”

Unlike new hiring or a reorganization of the graduate program, working toward a more supportive seminar culture is an action that individual faculty members can take on their own, without the need to create consensus at the departmental level. One idea is to save some comments on student work for a short private meeting held after the public seminar. Meeting one-on-one after the seminar can help students take stock of the various suggestions and prioritize which next steps are most important without the immediate pressures of a presentation.

We are agnostic about why some of the specific policies described here, including developing a more supportive seminar culture, might be particularly helpful for women students. Interview subjects mentioned the possibility that women are more averse to competitive environments, more likely to take harsh feedback personally, more likely to be left out of informal (“old boy”) networks, and so on. There is observational and experimental evidence for some of these channels (for example, see Goldin 2015 on differential response by gender to receiving low grades as undergraduates), and others channels would be interesting subjects for future study.

Awareness of Gender Bias

A final difference that we noticed between departments with better and worse relative outcomes for women was an awareness of gender bias, particularly in its more subtle and implicit forms, among the senior male faculty who often make up the majority of leadership positions in a department. Awareness of gender bias may influence how these senior faculty interact with women in their classes and in their advisory roles on a day-to-day basis. Moreover, we noticed that senior faculty with more cognizance of gender bias react differently to incidents of harassment (which, although rare, were mentioned at three of our five interview departments). One woman made this point quite succinctly, saying “my personal view is that a lot of it has to come from the men, [the field is] still very male dominated, so until the men are comfortable, nothing much good is going to happen for the women.”

Differences in awareness of gender bias were most obvious in responses to the interview question “do you see any difference in the [graduate department] environment for men and women?” Faculty at departments with better relative outcomes for women responded to the question in more observant and thoughtful ways. A senior faculty member at Department K was typical of this view in saying: “I’m confident that there is no explicit discrimination, but perhaps what is happening is more subtle or subconscious ... The same behaviors in a man say he’s forceful and defends his ideas; he’s aggressive in a good way. Those same behaviors when taken by women tend to get a different reaction in a subtle kind of way.”

In contrast, faculty members at Departments O and T focused exclusively on the lack of *overt* discrimination against women students and did not seem aware of (or to put much credence in) the subtle differences in the way men and women experience the culture of the field. A professor at Department O said “I’m not aware of people explicitly treating anyone different in one way or another [by gender], at least from faculty perspective. I can’t—I’m not aware of any specific instances to that effect. As far as I can tell, nothing that I can think of comes to mind.” A male faculty member from Department T answered similarly, reporting that he talked to graduate students and never heard complaints about poor treatment of women students. “I’m still a little perplexed about this,” he said. “I ask graduate students what the atmosphere is like for them, and I never hear about any problem of them being harassed in the kind of extreme form, but even in more subtle ways. My sense, and I acknowledge that I might be misreading, is that women in the department are treated well. That’s my sense, but I’m suspicious about if I’m gathering information correctly.”

Differences in awareness of gender issues can have important consequences in how department leadership respond to instances of harassment in a learning environment, which were mentioned at three of the five departments that we profiled. By harassment, we do not mean individual cases of unwanted sexual attention, although this may also happen. Instead, we heard about intentional attempts on the part of one or more men in the student body to make women graduate students feel unwelcome. These instances of harassment, while certainly not the norm, did occur across the board. We also noticed clear differences in how the faculty at each department responded to such events. In each case, administrators condemned the harassment and sought to punish the individuals responsible, but the responses differed in whether further action was taken to change the underlying culture or institutions that contributed to the problem.

At Department C, the faculty considered the event to be outside of the norm and immediately searched for an underlying cause, which they then acted to remedy. A faculty member at Department C tells this story: “We actually had an incident [recently] that I think was unusual, [details redacted]. One thing we realized is that we had [a new admissions director] for the last couple of years, and [...] we had ended up with a couple of classes with way too few women students. Maybe only 20 percent, rather than [our usual] much higher figure. And so, the chair basically leaned on the person in charge of admissions that we had to have a lot more women, and our entering class this year is 50–50, which is the first time it’s been so high. We’re hoping that means it won’t happen again, if you have more women even guys who think like that will feel pressured not to act. That was the first time that we had anything like that in all the years I’ve been here, and I was really pleased that the Chair really got on it and pressured the admissions chair rather than ignoring it.”

At Departments K and T, instead, the harassing behavior itself was condemned but there was no further action to assess and perhaps adjust the culture. A faculty member at Department T described an instance of harassment and the departmental response: “We’ve had problems in the past. For example, there were some men [engaging in redacted activities], which is an extreme form of harassment ... In

those cases, the chair went nuts. No one could figure out who was doing it, but they warned that they would be thrown out of the university if this happened again. I believe it has not happened anymore.” According to a student trained at Department K, a similar reaction occurred there in response to one individual’s harassing actions. “The Chair tried to make some statement the next year, held a meeting at the beginning of the year for all PhD students, and said that there had been [harassing behavior], people saying things about women that you shouldn’t say because you could hurt someone’s feelings.”

We acknowledge that improving awareness of gender bias—in oneself and one’s environment—is an amorphous recommendation, and hard to convert into direct action. Self-reflection, questioning one’s assumptions, and striving to learn more about the experiences of others are habits worth cultivating for any individual and especially academics. Empathy is a human virtue, and a scholarly one too. On an organizational level, attention and sensitivity to diverse perspectives in a department should be considered when making departmental leadership decisions.

Conclusions

We document substantial variation in the representation of women and in the early career success of women across economics PhD programs. We provide suggestive evidence, based on a set of structured interviews, that there are large differences in culture and practices across graduate programs in economics; these differences are associated with better and worse relative outcomes for women. Important differences across departments include the number of women on the faculty, regularized opportunities for contact between advisors and students, and collegial research seminars.

Of course, such qualitative observations are only the beginning of establishing hypotheses about “what works” in increasing the representation of women in economics. We encourage other scholars to test the hypotheses suggested by this paper in more detail. In addition, it would be interesting to know whether prospective graduate students consider the types of factors underlying these rankings when selecting departments. If so, the association between share of women in the graduate program and positive early career outcomes could be, in part, due to students’ choices in where to matriculate.

Typical studies of the underrepresentation of women in scientific and technical fields focus on early educational experiences (like differential treatment in high school courses by gender) or on the job market and in academic workplaces (for example, see Ceci, Ginther, Kahn, and Williams 2014). These emphases, while certainly important, overlook the potentially central role of the learning environment in graduate school itself. To the extent that the culture and policies of a graduate program matter, there may be meaningful steps that we can all take to encourage and support women in economics. We hope that our findings will help to inform ongoing discussions—or to initiate new ones—in departments seeking to ensure the best opportunities and outcomes for all their graduates.

■ We thank Leyla Mocan for conducting interviews and Bailey Palmer for excellent research assistance.

References

- Abrevaya, Jason, and Daniel S. Hamermesh.** 2012. "Charity and Favoritism in the Field: Are Female Economists Nicer (To Each Other)?" *Review of Economics and Statistics* 94(1): 202–207.
- Antecol, Heather, Kelly Bedard, and Jenna Stearns.** 2018. "Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?" *American Economic Review* 108(9): 2420–41.
- Bayer, Amanda, and Cecilia Elena Rouse.** 2016. "Diversity in the Economics Profession: A New Attack on an Old Problem." *Journal of Economic Perspectives* 30(4): 221–42.
- Blank, Rebecca M.** 1991. "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from *The American Economic Review*." *American Economic Review* 81(5): 1041–67.
- Bransch, Felix, and Michael Kvasnicka.** 2017. "Male Gatekeepers Gender Bias in the Publishing Process?" IZA Discussion Papers 11089, Institute for the Study of Labor (IZA).
- Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams.** 2014. "Women in Academic Science: A Changing Landscape." *Psychological Science in the Public Interest* 15(3): 75–141.
- Ginther, Donna K, and Shulamit Kahn.** 2004. "Women in Economics: Moving Up or Falling Off the Academic Career Ladder?" *Journal of Economic Perspectives* 18(3): 193–214.
- Ginther, Donna, and Shulamit Kahn.** 2014. "Academic Women's Careers in the Social Sciences." Chap. 11 in *The Economics of Economists: Institutional Setting, Individual Incentives, and Future Prospects*, edited by Alessandro Lanteri and Jack Vromen. Cambridge University Press.
- Goldin, Claudia.** 2015. "Gender and the Undergraduate Economics Major: Notes on the Undergraduate Economics Major at a Highly Selective Liberal Arts College." https://scholar.harvard.edu/files/goldin/files/claudia_gender_paper.pdf.
- Hale, Galina, and Tali Regev.** 2014. "Gender Ratios at Top PhD Programs in Economics." *Economics of Education Review* 41: 55–70.
- Hengel, Erin.** 2017. "Publishing While Female: Are Women Held to Higher Standards? Evidence from Peer Review." Cambridge Working Papers in Economics 1753, Faculty of Economics, University of Cambridge.
- Hilmer, Christiana, and Michael Hilmer.** 2007. "Women Helping Women, Men Helping Women? Same-Gender Mentoring, Initial Job Placements, and Early Career Publishing Success for Economics PhDs." *American Economic Review* 97(2): 422–26.
- Langan, Andrew.** 2018. "Female Managers and Gender Disparities: The Case of Academic Department Chairs." November, 16. https://scholar.princeton.edu/sites/default/files/alangan/files/langan_jmp_current.pdf.
- McDowell, John M., Larry D. Singell Jr, and Mark Stater.** 2006. "Two to Tango? Gender Differences in The Decisions to Publish and Coauthor." *Economic Inquiry* 44(1): 153–168.
- Neumark, David, and Rosella Gardecki.** 1998. "Women Helping Women? Role Model and Mentoring Effects on Female Ph.D. Students in Economics." *Journal of Human Resources* 33(1): 220–46.
- Sarsons, Heather.** 2017. "Recognition for Group Work: Gender Differences in Academia." *American Economic Review* 107(5): 141–45.

Fixing the Leaky Pipeline: Strategies for Making Economics Work for Women at Every Stage

Kasey Buckles

It is often said that the first step to solving a problem is to admit that you have one, and with regard to the representation of women in the economics profession, many economists now seem to have taken this step. Awareness of the gender gap in economics and interest in addressing it have been building slowly over recent years, but the issue was moved to the forefront by Alice Wu's (2018) study describing gender-biased language on a widely-read web forum for economists. Since its release as a working paper in the summer of 2017, numerous media features have described the discipline's "problem with women" (Coyle 2017; see also Gittleson 2017; Wolfers 2017, 2018; *The Economist* 2017). Against this backdrop, the American Economic Association (AEA) has hosted a high-profile session on Women in Economics at its annual meetings, formed a standing Committee on Equity, Diversity, and Professional Conduct, and adopted a Code of Professional Conduct (available at <https://www.aeaweb.org/about-aea/code-of-conduct>).

While the actions of the American Economic Association are welcome, lasting change will require effort at all levels of the profession. But what, exactly, can be done? We are fortunate in that, over the last decade, researchers have begun to evaluate strategies for attracting and retaining women in economics and other disciplines in which they are historically underrepresented. Some of this research has used the tools of rigorous policy evaluation, including randomized controlled trials,

■ *Kasey Buckles is the Brian and Jeannelle Brady Associate Professor of Economics, University of Notre Dame, Notre Dame, Indiana. She is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts, and a Research Fellow, Institute of Labor Economics, Bonn, Germany. Her email address is kbuckles@nd.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.33.1.43>

doi=10.1257/jep.33.1.43

to evaluate interventions. Thus, we now have high-quality evidence that can help identify approaches that are likely to be effective. In this paper, I will draw on this evidence to provide a toolkit for those who want to increase the representation of women in economics. I focus on smaller-scale, targeted interventions that might be implemented by individuals or small groups who want to try to “move the needle” in their spheres of influence—a professor teaching an undergraduate class, a mentor to graduate students, an active member of a professional organization, or a department chair looking to retain women on the faculty.

The experience of women economists in higher education has been characterized as a “leaky pipeline,” because the fraction of women in the discipline decreases at each stage along the path from graduate school to the full professor rank (*CSWEP News* 2018). The pipeline metaphor has been fairly criticized as being overly simplistic and linear (Branch 2016), but it remains useful for thinking about the issues women face and the possible solutions at different stages of academic life. Thus, in the sections that follow, I identify strategies for attracting and retaining undergraduate women, and for supporting the careers of graduate students and assistant and associate professors. Near the end of the paper, I turn my attention to the pipeline’s “source”: the K-12 experience. Finally, I consider ways that promising interventions could be adapted for women economists in nonacademic career paths, discuss broader policy changes that might also help close the gender gap in economics, and offer some guidance for future work.

Undergraduate Students

In recent years, women have comprised about 56 percent of undergraduate students in the United States, but less than one-third of economics majors. Moreover, the proportion of women majoring in economics has been flat for nearly 30 years, despite the rising share of women among undergraduates (Avilova and Goldin 2018).¹ Women are better represented in many similarly quantitative fields like math or the physical sciences than they are in economics.

Recognizing this issue, Tatyana Avilova and Claudia Goldin initiated the Undergraduate Women in Economics Challenge in 2015. Funded by a grant from the Alfred P. Sloan Foundation, the challenge aims to identify interventions that work to increase the numbers of women graduating with a degree in economics. The Challenge was implemented as a randomized controlled trial—specifically, 20 treatment schools were randomly chosen from a pool of 88 interested departments that had at least 30 economics graduates per year and were in the top 100 universities or colleges according to *U.S. News and World Report*. The treatment

¹In recent years, the percent women is slightly higher for senior economics majors than it is for freshmen declaring economics as a major, suggesting that the profession is having modest success in winning college women over to the discipline (*CSWEP News* 2018). I return to the issue of women’s choice of major as incoming college students in the section on K-12 students.

group received \$12,500 to spend on initiatives designed to increase the number of women in their major, targeting the incoming class of freshmen in the fall of 2015. Freshmen classes were targeted because “the prime moments where female students relative to male students decide to major in economics are at the very start of their undergraduate life and just after taking Principles” (Avilova and Goldin 2018, p. 4). Each department was allowed to design its own intervention as long as it did at least one of the following: 1) delivered better information to students, 2) provided mentoring or role models, or 3) altered instructional content and presentation style. These three objectives were chosen because they have been identified as particular challenges at the undergraduate level.

It is too soon to know the results of the Undergraduate Women in Economics Challenge. Students enrolling in fall 2015 are only now in their senior year, so the key outcome (economics graduates) is not yet observable. However, several treatment schools designed their own interventions as randomized controlled trials, creating “experiments within the experiment.”

For example, Colorado State chose an intervention in which sections of Principles courses were assigned to one of three groups (Li 2018). Students in sections in the first group received information about career prospects, average earnings, and grade distributions, and women whose grades were above the median at the midterm received an email encouraging them to major in economics.² Sections in the second group received this treatment, and all women were also invited to a series of mentoring activities. Sections in the third group (the control group) received no treatment. Women with midterm grades above the median who received either the full or partial treatment were about 6 percentage points more likely to major in economics, off a base of about 13 percent. These students were also more likely to say that they expected to enjoy studying and working in economics in a follow-up survey, and that they believed they could succeed in economics. There was no statistically significant difference in the effect of full or partial treatment, perhaps because take-up of the mentoring activities was very low.³ The intervention decreased men’s likelihood of majoring in economics—an effect driven by men with midterm grades below the median. Li shows that men overestimated the grade they would get in the class by more than one grade point on average (based on a pretreatment survey at the beginning of the semester), and hypothesizes that the information about the true grade distribution and their position in it may have discouraged some of these below-median men.

²Rask and Tiefenthaler (2008) find that women are less likely than men to continue in economics after receiving lower (but still good) grades in their Principles courses.

³While mentoring did not have an effect in the Colorado State program, the AEA’s Committee on the Status of Minority Groups in the Economics Profession conducts summer programs in which minority undergraduate students spend several weeks receiving both skills training and mentoring. Compared to students who applied to the program but did not attend, participants were more likely to enroll in and complete PhD programs in economics (Becker, Rouse, and Chen 2016). Among black economists working at academic institutions, those who participated in the program had better success with publications and grants (Price 2005).

As another example, Southern Methodist University (SMU), implemented a randomized controlled trial in which female graduates who majored in economics spoke to students in Principles classes about their careers (Porter and Serra 2018). The intervention was inspired by research showing that female role models—and in particular, female instructors—can influence women’s career choices. As Porter and Serra discuss in their careful review of this literature, most studies rely on correlations in the data, but a few have taken advantage of random assignment. Two of the most relevant studies exploit the random assignment of cadets at the US Air Force academy to introductory courses in science, technology, engineering, and mathematics fields. The results show that high-ability cadets who are assigned to a female instructor are more likely to major in, work in, and pursue advanced degrees in these fields (Carrell, Page, and West 2010; Mansour, Rees, Rintala, and Wozny 2018).

While the evidence on the effects of professor gender on female enrollment is compelling, in practice it may be difficult for departments to manipulate instructor gender. Many economics departments do not have enough female professors to ensure that all students experience a class with one of them (a “chicken and egg” problem). Placing a disproportionate share of women professors in classes that students take early in their studies, such as Principles classes, poses problems of its own.

The surprising and promising finding of the SMU study is that exposure to female role models *other than the instructor* can have a noticeable effect. In the study, sections of Principles classes were randomly chosen to be visited by two career women who had majored in economics at SMU. Untreated sections serve as the control group, and the authors also collected data on the (untreated) classes from the previous year, to add a differences-in-differences element to the study design. The intervention increased women’s enrollment in intermediate economics classes by 13 percentage points, and increased the probability that women expressed an intention to major in economics by 7.9 percentage points—roughly doubling both measures. The added students appear to have been drawn from those who otherwise would have majored in languages or the humanities, suggesting that this program’s success in attracting women to economics did not exacerbate problems in other fields in which women are underrepresented. Like the Colorado State study, the effects were largest for women with high grade point averages. The intervention had no effect on men’s outcomes.

Porter and Serra (2018, p. 24) acknowledge that the mechanisms for this large effect are unclear—did the intervention work through *information*, or through *inspiration*? If it is mostly information, then the information could possibly be provided in even more low-cost ways. The effect sizes in the SMU study are larger than those in the Colorado State study, which effectively provided only information and a nudge. As more results from the Undergraduate Women in Economics Challenge study become available, we may gain additional insight that will help identify the essential components of this effective treatment.⁴

⁴An alternative way to expose undergraduate women to female role models is to select a textbook that includes more women. Stevenson and Zlotnik (2018) surveyed eight commonly used Principles texts and found that the share of people referenced who are women ranged from 10 to 34 percent.

The Colorado State and SMU interventions focused on the first two areas targeted by the Undergraduate Women in Economics Challenge study—information and mentoring/role models. The third area is pedagogical innovation. The website “Div.E.Q.” (Diversifying Economic Quality at <http://diversifyingecon.org>), created by Amanda Bayer and sponsored by the Committee on the Status of Minority Groups in the Economics Profession, identifies several proven strategies that can be implemented in economics classrooms. For example, interactive learning techniques and values affirmation practices have been shown to eliminate gender gaps in introductory physics courses (Lorenzo, Crouch, and Mazur 2006; Miyake et al. 2010), with the latter having its largest effects on women who agreed with traditional stereotypes about gender and science. Courses that assess students using only high-pressure, timed exams are also more likely to result in anomalously low grades for women in introductory courses in science, technology, engineering, and mathematics classes (relative to their grade point average in other courses), suggesting that more diversified grading structures could help women succeed in those classes (Koester, Grom, and McKay 2016).

Graduate Students

The percent of students who are women in undergraduate and graduate programs in departments with doctoral programs has been nearly identical for the last several years, at around 30 percent; in departments without doctoral programs, the percent of undergraduates who are women is around 35 percent (*CSWEP News* 2018). Women matriculate from PhD programs at roughly the same rate at which they enter. The first significant leak in the academic pipeline occurs in the *transition* from graduate programs to assistant professorships. Of course, some of these women are moving into public or private sector jobs in economics. But given the research on the impact of female professors and role models cited in the previous section, it is important to ensure that women have a fair shot at making it through graduate school, and making the transition into an academic job if they wish to do so. In this section, I will discuss interventions that can improve the experience of women PhD students and help to ensure an even playing field for women on the academic job market for economists.

First, as with undergraduates, nudges that inform or encourage women have the potential to increase women’s attachment to their graduate program or their chances for success. As an example, Unkovic, Sen, and Quinn (2016) conducted a randomized controlled trial investigating the impact of personalized emails encouraging graduate students to submit papers to an applied statistics conference in the social sciences. Of the nearly 4,000 students in the experiment, half received the treatment and the other half received no emails. The intervention increased submissions to the conference by 2.7 percentage points, and the strongest effects were for women from top programs (4.3 percentage points). However, women in the treatment group were less likely to have their papers accepted than those in the

control. This might seem to suggest that the compliers were negatively selected, but in fact, treatment women who submitted to the conference were less likely to be in quantitative fields, and also were less likely to have a recommendation letter as part of the application. The authors suggest that female students lack networks and mentors that could help them navigate professional situations.⁵

When it comes to the job market, women can face biases that affect their probability of being hired. For example, a growing body of evidence suggests that students evaluate female instructors more harshly. MacNell, Driscoll, and Hunt (2015) found that students in an online-only course rated a course more highly when they were told that the instructor was male, regardless of the instructor's actual gender (see also Mengel, Sauermann, and Zölitz forthcoming). Women are also given less credit for coauthored papers (Sarsons 2017), and recommendation letters for women are less likely to refer to their ability or agency (Schmader, Whitehead, and Wysocki 2007; Madera, Hebl, and Martin 2009). While an audit study by Williams and Ceci (2015) showed that academics generally show a preference for hiring a woman over an identical male candidate, male economists were an exception in their hiring decisions. Furthermore, academics are *not* more likely to hire a woman over a slightly more qualified male (Ceci and Williams 2015). So, if biases in teaching evaluations or recommendation letters make an equally qualified woman appear slightly less qualified, women will be disadvantaged even when facing an unbiased hiring committee.

Some recent programs designed to reduce gender bias or to increase the number of women hires are promising. In a randomized controlled trial at the University of Wisconsin-Madison involving 92 academic departments in fields involving science, technology, engineering, and mathematics, half of the study's departments were randomly chosen to participate in a series of workshops on gender bias, while the other half served as a control group. The treatment group not only saw an increase in faculty members' awareness of gender bias issues and "self-efficacy to engage in gender-equity promoting behaviors" in the short term (Carnes et al. 2015), but also increased the proportion of women hired after the intervention by 18 percentage points (Devine et al. 2017). In departments where women were underrepresented, there was an increase in the probability of making a job offer to a woman (Fine et al. 2014). At Montana State University, search committees in science, technology, engineering, and math departments in a randomly selected treatment group received training and some resources to support the hiring of women, and "searches in the intervention were 6.3 times more likely to make an offer to a woman candidate, and women who were made an offer were 5.8 times more likely to accept" (Smith, Handley, Zale, Rushing, and Potvin 2015). At the University of Michigan, a faculty committee that focused on increasing the hiring of women met with

⁵The #EJMinfo twitter campaign began in 2017 as a way to make information about the job market and graduate school available to a wider audience, outside of traditional networks. In 2018, the American Economic Association introduced EconSpark, an online discussion forum with a similar aim (<https://www.aeaweb.org/economics-discussion-forum>).

departments and search committees; an analysis limited to a comparison of pre- and post-intervention outcomes found that hires of women in science and engineering more than doubled (Stewart, La Vaque-Manty, and Malley 2004).⁶ Soll, Milkman, and Payne (2015) discuss more strategies for reducing bias in evaluations.

I would be remiss if I did not raise an issue that can adversely affect women's experience in graduate school—sexual harassment. While harassment can occur at any stage of the pipeline, graduate students are especially vulnerable, as their career success is highly dependent on their relationships with advisors, mentors, and other students. In a 2015 survey of 27 institutions of higher learning, 44 percent of graduate and professional women students and 30 percent of men reported experiencing sexual harassment; of those, 16 percent of women and 11 percent of men say that the perpetrator was a teacher or advisor (Cantor et al. 2017).

Many universities have mandated sexual harassment training for faculty and staff, but these programs have not been rigorously evaluated. One exception is a study by Bingham and Scherer (2001), which randomized sexual harassment training at a large university at the department level. The treatment that the researchers are able to observe is whether the individual viewed the training videos. The findings indicate that effective sexual harassment training is difficult to implement and there is a potential for unintended consequences. In this case, faculty members who viewed the videos had a better understanding of policy issues surrounding sexual harassment. But perversely, men who viewed the videos expressed a lower willingness to report sexual harassment and were more likely to engage in victim blaming.

Many other approaches for decreasing sexual harassment have been offered—see the excellent CSWEP newsletter (2018) on the topic for examples. One specific proposal, which is also mentioned in the AEA's Ad Hoc Committee on the Professional Climate in Economics (2018), suggests an “information escrow” system (Ayers and Unkovic 2012). In this approach, an online platform such as Callisto (<https://www.projectcallisto.org/>) is set up that allows victims of sexual misconduct to file a time-stamped report. If anyone else identifies the same offender, the victims will be contacted by a counselor who will discuss options for additional action. This approach can encourage victims to come forward, knowing that they may be helping other victims or that their report may prevent future incidents. The report of the American Philosophical Association's Sexual Harassment Ad Hoc Committee (2013) contains further recommendations for curtailing sexual misconduct. But as sociologist Frank Dobbin stated in reference to the American Sociological Association's working group on harassment, “The research doesn't show a very clear solution here ... so we're at a point where we need organizations to try more things. And we need more research” (Parry 2018).

⁶In addition to these interventions specifically targeted at hiring women, more general efforts to make hiring more transparent or to standardize the evaluation of candidates can benefit traditionally underrepresented groups (Bragger, DeNicolis, Kutcher, Morgan, and Firth 2002; Uhlmann and Cohen 2005; Levashina, Hartwell, Morgeson, and Campion 2014). Efforts along these lines are currently underway at the Federal Reserve under the guidance of David Wilcox (2017).

Finally, a recent study by Bostwick and Weinberg (2018) uses quasi-random variation in cohort composition to show that women are more likely to persist in doctoral programs in science, technology, engineering, and mathematics fields when there are more women in their cohort. The effect is largest in male-dominated fields. If this result holds for economics, it implies that policies that increase the number of women in graduate programs could have a multiplier effect. That is, the marginal woman is not only more likely to earn a degree herself, but other women in her cohort are as well.

Assistant Professors

Once hired into the rank of assistant professor, women continue to face some of the same issues they faced as graduate students. But for many women, this stage also brings the additional pressures of the tenure or promotion process, increased family commitments, and the loss of formal mentoring structures. In this section, I will discuss strategies aimed at helping women succeed at this level so that their chances of being promoted to associate professor and receiving tenure are increased.

Given the importance of a strong publication record at this career stage, one strategy has been to reduce gender biases that could affect women's probabilities of getting a grant or having a paper accepted for publication. For example, a policy of "blind refereeing," which removes identifying information so that the reviewers of a paper or grant application do not know the identity or gender of the author, will also remove the opportunity to act on any biases such reviewers may hold. While laboratory-based studies have offered support for this idea (Knobloch-Westerwick, Glynn, and Huges 2013; Krawczyk and Smyk 2016), when this strategy has been evaluated in practice using natural experiments or audit studies, the results are mixed. In the late 1980s, the *American Economic Review* conducted a randomized controlled trial in which submitted papers were randomly assigned to either double- or single-blind review (Blank 1991). Both men and women fared worse under double-blind review; the effect was slightly stronger for men (suggesting that men benefit more from having their identities known), but the difference was not statistically significant. This finding is typical in this literature—double-blind journal review does seem to benefit women relative to men, but the effect size is small and not statistically significant in most studies (Tomkins, Zhang, and Heavlin 2017). Furthermore, there is no strong evidence of gender bias in reviewing in economics in the first place (Abrevaya and Hamermesh 2012; Bagues, Sylos-Labini, and Zinovyeva 2017). This research suggests that blind refereeing is unlikely to have a dramatic differential effect by gender. And in the internet era, it is increasingly difficult to conceal authors' identities anyway.

Another intervention that has been proposed to help women at this stage is targeted mentoring (for a survey, see Meschitti and Lawton Smith 2017). Mentoring can reduce information asymmetries, provide a support system, and connect women with role models (Croson and McGoldrick 2007). However, rigorous evidence on

the effects of mentoring has been hard to come by—people often self-select into mentoring relationships and programs, outcomes can be difficult to measure, and mentoring itself is difficult to define.

Within economics, we have evidence from a randomized controlled trial that addresses these challenges. The intervention is a two-day mentoring workshop sponsored by the AEA's Committee on the Status of Women in the Economics Profession and now held annually after the association's annual meetings. Begun in 2004, participants in the program (known as CeMENT) are organized into small groups by research field, with each group including both senior mentors and junior mentees. The groups offer feedback on one another's work. Participants also attend panel discussions on issues such as publishing, promotion, and work-life balance. Many groups continue to operate as support networks long after the workshop.

Initially, the workshops were funded in part by an ADVANCE grant from the National Science Foundation, and the grant included funds for a randomized control trial. A study evaluating intermediate outcomes for the first two cohorts of participants shows promising results (Blau, Currie, Croson, and Ginther 2010). Three years after the program, participants had 1.622 more publications and 0.09 more top-tier publications than the control sample on average. For the first cohort, outcomes were also observable five years after treatment, and the effect on total and top-tier publications was 2.677 and 0.200, respectively. The results are also supportive of a positive effect of the workshop on grant receipt. A follow-up study that expands the sample to later cohorts and examines the program's impact on promotion and retention rates is currently underway.⁷

Given this success, the CeMENT program has expanded over the years. Initially a biannual event, it is now held annually, with separate workshops for faculty from institutions with and without doctoral programs. Yet combined, the workshops still serve fewer than 100 women each year. How can the mentoring model be expanded or adapted so that more women can benefit? First, to increase scale, the intervention could be replicated at regional or field-specific conferences, or by a group of departments working collaboratively. Other disciplines have similar programs that could be used as models (Croson and McGoldrick 2007). Second, one of the aims of the workshop is to increase assistant professors' access to information about the profession; all of the handouts that CeMENT participants receive are now available online and can be shared. Third, department chairs or administrators can help junior women find mentors. A program at the University of Michigan (modeled on a similar program at Case Western University) provided new hires in science, technology, engineering, and mathematics fields with a "Launch Committee" of mentors. Participants had higher satisfaction with their service loads and with their work environment than similar nonparticipants (ADVANCE Program at the

⁷While not implemented as a randomized controlled trial, the NSF ADVANCE program also funded a mentoring program at Hunter College (Rabinowitz and Valian 2007). Assistant and associate professors who participated submitted more papers and grants in the two years after the program than they did in the year before.

University of Michigan 2015). Fourth, many institutions already have mentoring programs in place but do not provide much guidance to mentors; a randomized trial involving mentors of clinical and translational research scholars showed that an eight-hour mentor training program could improve mentees' perceptions of their mentor's skills and behaviors (Pfund et al. 2014).

Finally, recent research has highlighted a policy that actually works against women on the tenure track—"gender-neutral" clock stopping policies that allow both men and women to add time to the tenure clock with the birth of a child. While the policies are often adopted in the interest of fairness, they can disadvantage women if men are able to be more productive during their extended time due to differences in child-care responsibilities or the impact of the birth itself. Using data on hires into the top 50 economics departments between 1980 and 2005 and a model with gender-specific university fixed effects, Antecol, Bedard, and Stearns (2018) show that this is in fact the case: "men are 17 percentage points more likely to get tenure in their first job once there is an established gender-neutral clock stopping policy in place, while women are 19 percentage points less likely" (p. 2422). This seems to be due to the fact that men at institutions with gender-neutral clock stopping policies have more publications in top-five journals, perhaps because they take more risks with the additional time or have more time to develop papers. Chairs and departments could share this evidence with administrators and encourage them to think carefully about adopting these policies.

Associate Professors

The biggest "leak" in the career pipeline occurs as women move from the ranks of associate professor of economics to full or endowed professor. In 2017, about 23 percent of associate professors in departments with doctoral programs were women, compared to only 14 percent of full professors (*CSWEP News* 2018). This gap has actually widened since 1993, as the increase in the share of women among associate professors has outpaced that for full professors. Despite this, very few programs are targeted at helping women associate professors succeed, and those that do exist have not been carefully evaluated.

Nevertheless, some related research can guide the design of interventions at this stage. First, it seems likely that strategies that have been shown to help women at earlier stages in the pipeline could benefit tenured women as well—especially mentoring. Many women associate professors report a loss of mentoring and networks after tenure, perhaps because of "an assumption that tenured faculty have 'made it' and therefore can figure out the next steps with ease" (McQuillan, Holmes, Hill, and Anderson-Knott 2016, p. 71). But post-tenure challenges may be different, as Claire Potter (2013) describes in her article on the "associate professor blues," as family responsibilities change and expectations for service mount. As both of these issues have been shown to disproportionately affect women, mentoring that includes strategies for navigating them could help move more women into the

full professor ranks. A handful of universities including UNC-Charlotte, Brandeis University, Lehigh University, and Northeastern University (as discussed in Rochester Institute of Technology n.d.) have developed mid-career mentoring programs. Also, CSWEP now hosts a mentoring breakfast for mid-career women at the annual AEA meetings in addition to its breakfast for junior women.

Second, women in academia do more service, and the service they do is more likely to be internal (and therefore less likely to be rewarded) (Guarino and Borden 2017). Babcock, Recalde, Vesterlund, and Weingart (2017) conducted a clever laboratory experiment on “low promotability” tasks, in which groups were asked to push a button. All group members received a reward if someone pushed the button, but the person who volunteered to push it received a lower reward. In mixed-gender groups, women were 50 percent more likely to push the button. When one group member was tasked with asking someone to volunteer, they were more likely to ask a woman. This reveals a dangerous cycle—women are more likely to be asked to volunteer, and are more likely to say “yes,” which reinforces incentives to ask them.

To address this service gap, department chairs and university administrators could develop clear service expectations and systems for quantifying and weighting service contributions so that it is easier to see when service work is not being equally shared. Such a system will also guide faculty as they prepare for promotion and should be used by promotion committees in their deliberations. Chairs could also develop service plans for all faculty members, to ensure that these responsibilities are being distributed evenly. A service plan also makes it easier for a faculty member to say “no” when appropriate, because she will have a directive from her chair.

Third, at most institutions, the criteria for promotion to full professor include some degree of visibility within the profession at large. This can be measured by invitations to participate in panels, to give a keynote address, or to serve on editorial boards. When these invitations are issued through existing networks, they can inadvertently exclude qualified women (resulting in the dreaded “manel”). Departments and professional organizations could have policies on representation, and could recommend best practices. In political science, the University of Arizona School of Government and Policy has sponsored an online, searchable database named “Women Also Know Stuff” (at <https://womenalsoknowstuff.com/>) on which women can register and identify their areas of expertise. The site is for academics and journalists to use “when writing syllabi; when planning conferences, panels, and speaker series; when citing research; when inviting essays and op-eds; and when identifying experts for articles.” This model has recently been adopted in history, philosophy, and other science, technology, engineering, and mathematics fields. A similar database in economics could be developed by a professional organization or department.

K-12 Students

The career pipeline as I have described it thus far begins at the undergraduate level, as this is the point at which most academics first encounter potential

economists. However, a serious discussion of strategies for closing the gender gap in economics must also include a look at the pipeline's source—the K-12 level. Large gender gaps in college major intentions among incoming students suggest that many women are being discouraged from studying economics before they ever enter a Principles classroom (Goldin 2015). Avilova and Goldin (2018) offer an explanation: “Students often think that economics is only for those who want to work in the financial and corporate sectors and do not realize that economics is also for those with intellectual, policy and career interests in a wide range of fields” (p. 1). If women are less interested in finance and business (putting aside how those preferences are formed), then we could be losing many potential economists right out of the gate as a result of this misperception.⁸

One possible area for improvement is the design of Advanced Placement (AP) exams and courses, which are developed by the College Board in collaboration with academic economists. Because most AP exam takers are college bound, this is an ideal population to target for an intervention. Economists could work with the College Board to ensure that the AP courses present a more complete picture of what economists do. For example, the current AP Microeconomics course focuses on product markets, factor markets, and market failures, and the role of government (College Board AP 2012). While these topics should remain central, the course could shift its content to include discussions of how economists apply these concepts to topics like health, education, family, crime, or development. These fields are relatively popular among women academics (Chari and Goldsmith-Pinkham 2017); it would not be surprising if they were also more appealing to women in high school.

A change to the AP economics curriculum has the potential to affect many students. In 2017, 141,649 students sat for the Macroeconomics exam and 87,858 sat for the Microeconomics exam; 44 percent of the exams were taken by women.⁹ Moreover, if these changes were to spill over into the broader high school economics curriculum, the impact could be even larger—22 states currently require an economics course for graduation, and sixteen require standardized testing of economic concepts (Council for Economic Education 2018).

Fricke, Grogger, and Steinmayr (2018) provide evidence suggesting that this approach could increase women's likelihood of choosing to study economics. The authors exploit a natural experiment in which first-year college students are required to write a paper on a topic that is assigned quasi-randomly. Students assigned to write in economics were more likely to choose it as a major, and the effect was largest when students were assigned a paper in an area “*less typical of the public's perception of the field of economics*” (p. 199, emphasis added). Wang and Degol (2017) suggest that teaching adolescent girls about the ways that innovations in science

⁸When my daughter was in second grade, her teacher began a unit on economics by asking if anyone knew what economists study. My daughter, having heard me talk about my research, raised her hand and answered “families.” Her teacher kindly told her that she was incorrect, and that economists study money.

⁹As reported at College Board, “Student Score Distributions, AP Exams—May 2017,” <https://secure-media.collegeboard.org/digitalServices/pdf/research/2017/Student-Score-Distributions-2017.pdf>.

and engineering improve people's quality of life could help close stubborn gender gaps across the science, technology, engineering, and mathematics fields. Indeed, some universities that have developed minors or programs that emphasize the social impact of the work of engineers have managed to eliminate or even reverse those gaps (Nilsson 2015).

Efforts to broaden students' understanding of economics fall outside the range of small, carefully evaluated interventions that I have highlighted for the other stages of the pipeline. I include this discussion because it is unlikely that economists will make substantial and lasting progress toward gender balance if we ignore the K-12 experience. More innovation and research is needed on this front; see the National Research Council's (2006) book "To Recruit and Advance" for strategies that academics can use to reach K-12 students.

Discussion

The set of potential strategies for closing the gender gap in economics is large. In this paper, I have limited the set in three ways. First, I have framed the discussion of strategies for retaining women in academic careers along the tenure track. However, nearly all of these interventions could also be adapted for women in the public or private sectors, or in non-tenure-track academic positions. For example, most jobs have a "service" component that may disproportionately fall on women and especially women of color (Cross, Rebele, and Grant 2016; Williams and Multhaup 2018), so many women could benefit from systems that measure and reward this work. Mentoring programs like the CeMENT workshops described earlier could be adapted for women in other career paths and implemented by professional organizations, or by firms and government agencies that employ women economists. Likewise, many of these strategies could be adapted to help increase the numbers of economists from other underrepresented groups; see the wiki on "Div.E.Q." for examples of research in this area (at <http://diversifyingecon.org>).

Second, I have focused on smaller-scale, targeted interventions that could be implemented by individuals, organizations, or academic units who are working to attract and retain women students and faculty. Of course, some public policies could also help to accomplish this goal by making it easier for all women to manage professional and family responsibilities. For example, maternity and family leave policies have been shown to improve job continuity and to have long-lasting positive effects on employment, though leaves longer than a year may negatively impact women's careers (Hegewisch and Gornick 2011; Rossin-Slater 2018). Better access to child care can help women maintain employment and work more hours in demanding careers (Furtado and Hock 2010; Cortes and Tessada 2011; Hegewisch and Gornick 2011), while flexible work arrangements are associated with reduced work-family conflict and increased job attachment (Gajendran and Harrison 2007). Though such policies are beyond the scope of this paper, they will likely be important for sustained progress toward narrowing the gender gap in economics.

Third, to the extent possible, I have discussed interventions that have been evaluated in a way that allows for credible estimation of a causal effect. However, those willing to work toward greater diversity and inclusion in economics should not ignore the wealth of resources that describe thoughtful and inventive strategies that do not (yet) meet this standard (for examples, see National Research Council 2006; Stewart, Malley, and LaVaque-Manty 2007; *CSWEP News* 2017; Stewart and Valian 2018). Research in this area faces several challenges to causal identification—it can be difficult to find reasonable comparison samples, treatments are often bundled, and outcomes can be difficult to measure or take years to be realized. Even in cases where a “gold standard” randomized trial was used to evaluate an intervention, questions about external validity and replicability remain. As a result, we will likely have to try some things without knowing for certain that they will work. This is especially true at the associate professor and K-12 levels, where rigorous evidence on effective strategies is virtually nonexistent. But this challenge also presents an opportunity for the profession to put its policy evaluation toolkit to work to advance knowledge, by building evaluation into implementation of policies or by finding creative ways to evaluate interventions after they have occurred. To aid this effort, the AEA could create a registry for randomized controlled trials aimed at attracting or retaining women and other underrepresented groups to economics.¹⁰

Finally, while many of the interventions discussed here have been shown to yield important benefits, they also have costs. The CeMENT mentoring program requires a significant time investment from its mentors and even more from its organizers; anti-bias training can be expensive and requires faculty time; developing service plans and transparent systems for rewarding service can be complex and (again) time-consuming. Even lower-cost interventions like having successful alumni speak to Principles classes require time to coordinate and manage. As Bayer and Rouse (2016) argued in this journal, the entire economics profession is likely to benefit from a more diverse membership, as the range of views that are represented expands and group dynamics and decision-making improve. The costs of working toward that goal should not fall on the shoulders of women and underrepresented groups alone.

■ *I would like to thank Lisa Argyle, Mary Flannery, Claudia Goldin, Nora Gordon, Gary Hoover, Laura Kramer, Elira Kuka, Stephanie Rennane, Claudia Sahn, Rhonda Sharpe, and Abigail Wozniak for helpful comments and conversations. I am also grateful to Alison Doxey for her superb research assistance.*

¹⁰Wang and Degol (2017) identify several questions for future research on strategies for closing gender gaps in STEM fields.

References

- Abrevaya, Jason, and Daniel S. Hamermesh. 2012. "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?" *Review of Economics and Statistics* 94(1): 202–207.
- ADVANCE Program at the University of Michigan. 2015. "Launch Committee Program Summary Report of Launchee Data." October. <https://advance.umich.edu/wp-content/uploads/2018/09/LaunchreportFinal.pdf>.
- AEA Ad Hoc Committee on the Professional Climate in Economics. 2018. Interim Report, April 6. <https://www.aeaweb.org/resources/member-docs/report-on-professional-climate>.
- American Philosophical Association Sexual Harassment Ad Hoc Committee. 2013. "We Can Act: A Report Recommending Best Practices to be Implemented by the APA regarding Sexual Harassment in the Discipline." Submitted for discussion at the November 2013 National APA Board Meeting. <https://cdn.ymaws.com/www.apaonline.org/resource/resmgr/sexualharassmentreport.pdf>.
- Antecol, Heather, Kelly Bedard, and Jenna Stearns. 2018. "Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?" *American Economic Review* 108(9): 2420–41.
- Avilova, Tatyana, and Claudia Goldin. 2018. "What Can UWE Do for Economics?" NBER Working Paper 24189.
- Ayres, Ian, and Cait Unkovic. 2012. "Information Escrows." *Michigan Law Review* 111(2): 145–96.
- Babcock, Linda, Maria P. Recalde, Lise Vesterlund, and Laurie Weingart. 2017. "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability." *American Economic Review* 107(3): 714–47.
- Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva. 2017. "Does the Gender Composition of Scientific Committees Matter?" *American Economic Review* 107(4): 1207–38.
- Bayer, Amanda, and Cecilia Elena Rouse. 2016. "Diversity in the Economics Profession: A New Attack on an Old Problem." *Journal of Economic Perspectives* 30(4): 221–42.
- Becker, Charles M., Cecilia Elena Rouse, and Mingyu Chen. 2016. "Can a Summer Make a Difference? The Impact of the American Economic Association Summer Program on Minority Student Outcomes." *Economics of Education Review* 53: 46–71.
- Bingham, Shereen G., and Lisa L. Scherer. 2001. "The Unexpected Effects of a Sexual Harassment Educational Program." *Journal of Applied Behavioral Science* 37(2): 125–53.
- Blank, Rebecca M. 1991. "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from *The American Economic Review*." *American Economic Review* 81(5): 1041–1067.
- Blau, Francine D., Janet M. Currie, Rachel T. A. Croson, and Donna K. Ginther. 2010. "Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial." *American Economic Review* 100(2): 348–52.
- Bostwick, Valerie K., and Bruce A. Weinberg. 2018. "Nevertheless She Persisted? Gender Peer Effects in Doctoral STEM Programs." NBER Working Paper 25028.
- Bragger, Jennifer DeNicolis, Eugene Kutcher, John Morgan, and Patricia Firth. 2002. "The Effects of the Structured Interview on Reducing Biases Against Pregnant Job Applicants." *Sex Roles* 46(7–8): 215–26.
- Branch, Enobong Hannah. 2016. *Pathways, Potholes, and the Persistence of Women in Science: Reconsidering the Pipeline*. London: Lexington Books.
- Cantor, David, Bonnie Fisher, Susan Chibnall, Reanne Townsend, Hyunshik Lee, Carol Bruce, and Gail Thomas. 2017. *Report on the AAU Campus Climate Survey on Sexual Assault and Sexual Misconduct*. Rockville, MD: Westat.
- Carnes, Molly, Patricia G. Devine, Linda Baier Manwell, Angela Byars-Winston, Eve Fine, Cecilia E. Ford, Patrick Forscher, Carol Issac, Anna Kaatz, Wairimu Magua, Mari Palta, and Jennifer Sheridan. 2015. "The Effect of an Intervention to Break the Gender Bias Habit for Faculty at One Institution: A Cluster Randomized, Controlled Trial." *Academic Medicine: Journal of the Association of American Medical Colleges* 90(2): 221–30.
- Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics* 125(3): 1101–144.
- Ceci, Stephen J., and Wendy M. Williams. 2015. "Women have Substantial Advantage in STEM Faculty Hiring, Except when Competing Against More-Accomplished Men." *Frontiers in Psychology* 6: 1532.
- Chari, Anusha, and Paul Goldsmith-Pinkham. 2017. "Gender Representation in Economics Across Topics and Time: Evidence from the NBER Summer Institute." NBER Working Paper 23953.
- College Board AP. 2012. "Economics: Macroeconomics, Microeconomics: Course Description, Effective Fall 2012." <http://media.collegeboard.com/digitalServices/pdf/ap/ap-economics-course-description.pdf>.

- Cortés, Patricia, and José Tessada.** 2011. "Low-Skilled Immigration and the Labor Supply of Highly Skilled Women." *American Economic Journal: Applied Economics* 3(3): 88–123.
- Council for Economic Education.** 2018. "2018 Survey of the States: Economics and Personal Finance Education in Our Nation's Schools." Available at: <https://www.councilforeconed.org/policy-and-advocacy/survey-of-the-states-2/>.
- Coyle, Diane.** 2017. "Economics has a Problem with Women." *Financial Times*, August 28.
- Crosen, Rachel, and KimMarie McGoldrick.** 2007. "Scaling the Wall: Helping Female Faculty in Economics Achieve Tenure." Chap. 10 in *Transforming Science and Engineering: Advancing Academic Women*, edited by Abigail J. Stewart, Janet E. Malley, and Danielle LaVaque-Manty, 152–69. University of Michigan Press.
- Cross, Rob, Reb Rebele, and Adam Grant.** 2016. "Collaborative Overload." *Harvard Business Review* January–February, 94(1): 16.
- CSWEP News.** 2017. Issue II. <https://www.aeaweb.org/content/file?id=5524>.
- CSWEP News.** 2018. Issue I. <https://www.aeaweb.org/content/file?id=6769>
- Devine, Patricia G., Patrick S. Forscher, William T. L. Cox, Anna Kaatz, Jennifer Sheridan, and Molly Carnes.** 2017. "A Gender Bias Habit-Breaking Intervention Led to Increased Hiring of Female Faculty in STEM Departments." *Journal of Experimental Social Psychology* 73: 211–15.
- Fine, Eve, Jennifer Sheridan, Molly Carnes, Jo Handelsman, Christine Pribbenow, Julia Savoy, and Amy Wendt.** 2014. "Minimizing the Influence of Gender Bias on the Faculty Search Process." In *Gender Transformation in the Academy*, edited by Vasiliki Demos, Catherine White Berheide, and Marcia Texler Segal, 267–89. UK: Emerald Group.
- Fricke, Hans, Jeffery Grogger, and Andreas Steinmayr.** 2018. "Exposure to Academic Fields and College Major Choice." *Economics of Education Review* 64: 199–213.
- Furtado, Delia, and Heinrich Hock.** 2010. "Low Skilled Immigration and Work–Fertility Tradeoffs among High Skilled US Natives." *American Economic Review* 100(2): 224–28.
- Gajendran, Ravi S., and David A. Harrison.** 2007. "The Good, the Bad, and the Unknown about Telecommuting: Meta-Analysis of Psychological Mediators and Individual Consequences." *Journal of Applied Psychology* 92(6): 1524–541.
- Gitlerson, Kim.** 2017. "Where Are All the Women in Economics?" *BBC News*, October 13, 2017.
- Goldin, Claudia.** 2015. "Gender and the Undergraduate Economics Major: Notes on the Undergraduate Economics Major at a Highly Selective Liberal Arts College." https://scholar.harvard.edu/files/goldin/files/claudia_gender_paper.pdf.
- Guarino, Cassandra M., and Victor M. Borden.** 2017. "Faculty Service Loads and Gender: Are Women Taking Care of the Academic Family?" *Research in Higher Education* 58(6): 672–94.
- Hegewisch, Ariane, and Janet C. Gornick.** 2011. "The Impact of Work-Family Policies on Women's Employment: A Review of Research from OECD Countries." *Community, Work & Family* 14(2): 119–38.
- Economist, The.** 2017. "Inefficient Equilibrium: Women in Economics." December 19. <https://www.economist.com/christmas-specials/2017/12/19/women-and-economics>.
- Knobloch-Westerwick, Silvia, Carroll J. Glynn, and Michael Huge.** 2013. "The Matilda Effect in Science Communication: An Experiment on Gender Bias in Publication Quality Perceptions and Collaboration Interest." *Science Communication* 35(5): 603–25.
- Koester, Benjamin P., Galina Grom, and Timothy A. McKay.** 2016. "Patterns of Gendered Performance Difference in Introductory STEM Courses." arXiv: 1608.07565.
- Krawczyk, Michał, and Magdalena Smyk.** 2016. "Author's Gender Affects Rating of Academic Articles: Evidence from an Incentivized, Deception-Free Laboratory Experiment." *European Economic Review* 90: 326–35.
- Levashina, Julia, Christopher J. Hartwell, Frederick P. Morgeson, and Michael A. Campion.** 2014. "The Structured Employment Interview: Narrative and Quantitative Review of the Research Literature." *Personnel Psychology* 67(1): 241–93.
- Li, Hsueh-Hsiang.** 2018. "Do Mentoring, Information, and Nudge Reduce the Gender Gap in Economics Majors?" *Economics of Education Review* 64: 165–83.
- Lorenzo, Mercedes, Catherine H. Crouch, and Eric Mazur.** 2006. "Reducing the Gender Gap in the Physics Classroom." *American Journal of Physics* 74(2): 118–22.
- MacNell, Lillian, Adam Driscoll, and Andrea N. Hunt.** 2015. "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching." *Innovative Higher Education* 40(4): 291–303.
- Madera, Juan M., Michelle R. Hebl, and Randi C. Martin.** 2009. "Gender and Letters of Recommendation for Academia: Agent and Communal Differences." *Journal of Applied Psychology* 94(6): 1591–99.
- Mansour, Hani, Daniel I. Rees, Bryson Rintala, and Nathan Wozny.** 2018. "The Effects of Professor Gender on the Post-Graduation Outcomes of Female Students." IZA Discussion Paper 11820,

Institute for the Study of Labor (IZA).

McQuillan, Julia, Mary Anne Holmes, Patricia Wonch Hill, and Mindy Anderson-Knott. 2016. "Does the Road Improve in the Land of the Tenured? Exploring Perceptions of Culture and Satisfaction by Rank and Gender." Chap. 4 in *Pathways, Potholes, and the Persistence of Women in Science: Reconsidering the Pipeline*, edited by Enobong Hannah Branch. London: Lexington Books.

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. Forthcoming. "Gender Bias in Teaching Evaluations." *Journal of the European Economic Association*.

Meschitti, Viviana, and Helen Lawton Smith. 2017. "Does Mentoring Make a Difference for Women Academics? Evidence from the Literature and a Guide for Future Research." *Journal of Research in Gender Studies* 7(1): 166–99.

Miyake, Akira, Lauren E. Kost-Smith, Noah D. Finkelstein, Steven J. Pollock, Geoffrey L. Cohen, and Tiffany A. Ito. 2010. "Reducing the Gender Achievement Gap in College Science: A Classroom Study of Values Affirmation." *Science* 330(6008): 1234–37.

National Research Council. 2006. *To Recruit and Advance: Women Students and Faculty Science and Engineering*. Washington, DC: National Academies Press. <https://doi.org/10.17226/11624>.

Nilsson, Lina. 2015. "How to Attract Female Engineers." *New York Times*, April 27.

Parry, Marc. 2018. "In Sociologists' #MeToo Moment, Does Their Response Heed the Lessons of Social Science?" *Chronicle of Higher Education*, September 10, 2018.

Pfund, Christine, Stephanie C. House, Pamela Asquith, Michael F. Fleming, Kevin A. Buhr, Ellen L. Burnham, Julie M. Eichenberger Gilmore, W. Charles Huskins, Richard McGee, Kathryn Schurr, Eugene D. Shapiro, Kimberly C. Spencer, and Christine A. Sorkness. 2014. "Training Mentors of Clinical and Translational Research Scholars: A Randomized Controlled Trial." *Academic Medicine: Journal of the Association of American Medical Colleges* 89(5): 774–82.

Porter, Catherine, and Danila Serra. 2018. "Gender Differences in the Choice of Major: The Importance of Female Role Models." April 12.

Potter, Claire. 2013. "The Associate Professor Blues." *Chronicle of Higher Education*, September 28.

Price, Gregory N. 2005. "The Causal Effects of Participation in the American Economic Association Summer Minority Program." *Southern Economic Journal* 72(1): 78–97.

Rabinowitz, Vita C., and Virginia Valian. 2007. "Beyond Mentoring: A Sponsorship Program to Improve Women's Success." Chap. 7 in *Transforming Science and Engineering: Advancing Academic*

Women, edited by Abigail J. Stewart, Janet E. Malley, and Danielle LaVaque-Manty. University of Michigan Press.

Rask, Kevin, and Jill Tiefenthaler. 2008. "The Role of Grade Sensitivity in Explaining the Gender Imbalance in Undergraduate Economics." *Economics of Education Review* 27(6): 676–87.

Rochester Institute of Technology. No date. "Mentoring for Mid-Career Faculty." <https://www.rit.edu/academicaffairs/facultydevelopment/faculty-mentoring/mentoring-mid-career-faculty> (webpage accessed September 5, 2018).

Rossin-Slater, Maya. 2018. "Maternity and Family Leave Policy." In *The Oxford Handbook of Women and the Economy*, edited by Susan L. Averett, Laura M. Argys, and Saul D. Hoffman, 323–342. New York: Oxford University Press.

Sarsons, Heather. 2017. "Recognition for Group Work: Gender Differences in Academia." *American Economic Review* 107(5): 141–45.

Schmader, Toni, Jessica Whitehead, and Vicki H. Wysocki. 2007. "A Linguistic Comparison of Letters of Recommendation for Male and Female Chemistry and Biochemistry Job Applicants." *Sex Roles* 57(7–8): 509–14.

Smith, Jessi L., Ian M. Handley, Alexander V. Zale, Sara Rushing, and Martha A. Potvin. 2015. "Now Hiring! Empirically Testing a Three-Step Intervention to Increase Faculty Gender Diversity in STEM." *BioScience* 65(11): 1084–87.

Soll, Jack B., Katherine L. Milkman, and John W. Payne. 2015. "A User's Guide to Debiasing." Chap. 33 in *The Wiley Blackwell Handbook of Judgment and Decision Making*, vol. 2, edited by Gideon Keren and George Wu. UK: John Wiley & Sons, Ltd.

Stevenson, Betsey, and Hanna Zlotnik. 2018. "Representations of Men and Women in Introductory Economics Textbooks." *AEA Papers and Proceedings* 108: 180–85.

Stewart, Abigail J., Danielle La Vaque-Manty, and Janet E. Malley. 2004. "Recruiting Female Faculty Members in Science and Engineering: Preliminary Evaluation of One Intervention Model." *Journal of Women and Minorities in Science and Engineering* 10(4): 361–75.

Stewart, Abigail J., Janet E. Malley, and Danielle LaVaque-Manty, eds. 2007. *Transforming Science and Engineering: Advancing Academic Women*. University of Michigan Press.

Stewart, Abigail J., and Virginia Valian. 2018. *An Inclusive Academy: Achieving Diversity and Excellence*. Cambridge, MA: The MIT Press.

Tomkins, Andrew, Min Zhang, and William D. Heavlin. 2017. "Single versus Double Blind Reviewing at WSDM 2017." arXiv: 1702.00502.

Uhlmann, Eric Luis, and Geoffrey L. Cohen.

2005. "Constructed Criteria: Redefining Merit to Justify Discrimination." *Psychological Science* 16(6): 474–80.

Unkovic, Cait, Maya Sen, and Kevin M. Quinn. 2016. "Does Encouragement Matter in Improving Gender Imbalances in Technical Fields? Evidence from a Randomized Controlled Trial." *PLoS ONE* 11(4): e0151714.

Wang, Ming-Te, and Jessica L. Degol. 2017. "Gender Gap in Science, Technology, Engineering, and Mathematics (STEM): Current Knowledge, Implications for Practice, Policy, and Future Directions." *Educational Psychology Review* 29(1): 119–40.

Wilcox, David. 2017. "Diversity and Inclusion at the Federal Reserve Board." *CSWEP News*, Issue II, pp. 5–9. <https://www.aeaweb.org/content/>

file?id=5524.

Williams, Joan C., and Marina Multhaup. 2018. "For Women and Minorities to Get Ahead, Managers Must Assign Work Fairly." *Harvard Business Review*, March 5.

Williams, Wendy M., and Stephen J. Ceci. 2015. "National Hiring Experiments Reveal 2:1 Faculty Preference for Women on STEM Tenure Track." *PNAS* 112(17): 5360–65.

Wolfers, Justin. 2017. "Evidence of a Toxic Environment for Women in Economics." *New York Times*, August 18.

Wolfers, Justin. 2018. "Why Women's Voices are Scarce in Economics." *New York Times*, February 2.

Wu, Alice H. 2018. "Gendered Language on the Economics Job Market Rumors Forum." *AEA Papers and Proceedings* 108: 175–79.

Financial Regulation: Still Unsettled a Decade After the Crisis

Daniel K. Tarullo

A decade after the darkest moments of the financial crisis, both the US financial system and the legal framework for its regulation are still in flux. The post-crisis regulatory framework has made systemically important banks much more resilient. They are substantially better capitalized and less dependent on runnable short-term funding. But the current regulatory framework does not deal effectively with threats to financial stability outside the perimeter of regulated banking organizations, notably from forms of shadow banking. Moreover, with the political tide having for the moment turned decisively toward deregulation, there is some question whether the resiliency improvements of the largest banks will be preserved.

This article assesses the accomplishments, unfinished business, and outstanding issues in the post-crisis approach to prudential regulation. After briefly reviewing how the ongoing integration of capital markets and traditional lending channels undermined the New Deal regulatory framework, I explain how the post-crisis regulatory approach of instigating changes across a range of bank activities and practices brought about a steady improvement in the resiliency of the financial system, especially in the largest financial institutions. Next, I turn to an evaluation of how

■ *Daniel K. Tarullo is Professor of Law, Georgetown University, Washington, DC, and Visiting Professor of Law, Harvard Law School, Cambridge, Massachusetts. From 2009 to 2017, he was a member of the Board of Governors of the Federal Reserve, Washington, DC, where he served as oversight Governor for Supervision and Regulation. His email address is dtarullo@law.harvard.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at

<https://doi.org/10.1257/jep.33.1.61>

doi=10.1257/jep.33.1.61

durable this regulatory approach will prove over time. The answer will depend on how financial regulators can and will respond to what will surely be the highly adaptive behavior of financial market participants to changes in regulation, technology, and the overall market environment. The hurdles to doing so, both political and institutional, are substantial. While regulators have ample legal authority to contain risks at prudentially regulated banking organizations, over time they may lack the will or organizational capacity to exercise those authorities effectively. It is doubtful whether they have adequate authority to address threats to financial stability that may arise *outside* the perimeter of prudentially regulated firms. In particular, there is reason for concern about appropriate regulation of liquidity and short-term finance, which would likely be at the center of a future crisis. Thus, while the resiliency of the financial system is likely to remain fairly high in the near term, the medium- and longer-term prospects are hazier than one might hope.¹

A Brief Overview of Financial Regulation Since the New Deal

The Banking Act of 1933, more commonly known as the Glass–Steagall Act, which passed against the backdrop of a different kind of financial crisis, adopted a structural approach by separating and then, in various ways, protecting commercial banking from investment banking and trading. The ensuing stability in the banking system occasioned only modest changes in financial regulation for the better part of 40 years. However, from the late 1960s into the early 1970s, the demise of the Bretton Woods system, adverse macroeconomic conditions, and increasing international competition posed big challenges for what had been the relatively safe and profitable—but not particularly innovative—business of commercial banking. In addition, capital markets activities made incursions into areas previously dominated by commercial banks. As inflation ballooned in the 1970s, banks faced regulatory limits on the interest rate they could pay on bank deposits (“Regulation Q”), and funds flowed into money markets funds instead. On the asset side of the balance sheet, public corporations increasingly turned to commercial paper and public debt markets, thus reducing their demand for bank loans.

As bankers faced the erosion of barriers to competition from nonbank financial institutions, they sought regulatory relief on everything from limits on interstate branching to restrictions on their affiliating with securities underwriters. Bank regulators and legislators, fearing a continued decline of the franchise value of traditional depository institutions, provided a good bit of that relief over a quarter century: for example, the Riegle–Neal Interstate Banking and Branching Efficiency Act of 1994 allowed banks to branch across state lines, and the Financial Services Modernization Act of 1999 (the Gramm–Leach–Bliley Act) allowed commercial

¹This essay will not discuss a number of important nonprudential regulatory topics covering the financial system, including cybersecurity, consumer protection, and investor protection.

banks, investment banks, securities firms, and insurance companies to combine under the umbrella of a bank holding company.

Given that market pressures had undermined the rationale for, and effectiveness of, the New Deal regulatory framework, many of these steps were reasonable. However, the old regulatory framework was not replaced with a new one, other than the formalization of relatively modest capital requirements following the troubles of Latin American debt and the savings and loan industry in the 1980s. The removal of many activity and affiliation restrictions freed banks to grow and to acquire or develop trading and investment banking units. Meanwhile, “shadow banking” grew rapidly: the term refers to nonbank financial institutions, including investment banks and money market funds, that receive short-term funding and then make loans or invest in debt-related assets.

This progressive integration of capital markets and traditional lending played a major role in the unsustainable explosion of subprime lending and mortgage-backed securities in the 2000–2006 period. A sharp reduction in the ability of shadow banks to raise capital in short-run markets has an effect similar to the bank runs in the early 1930s, before federal deposit insurance was established. In the 2007–2009 crisis, these new vulnerabilities manifested themselves with a vengeance when short-term wholesale funding (such as repurchase agreements, or “repos”) dried up, as investors reacted to the decline in housing prices and questions about the underwriting of mortgage-backed securities.

While the Dodd–Frank Wall Street Reform and Consumer Protection Act that passed in July 2010 emphasized financial stability and systemic risk, it largely eschewed the structural regulatory solutions of the 1930s. It did not seek to reinstate the Glass–Steagall separation of commercial and investment banking, nor to break larger banks into smaller sizes. Instead, the law enacted a broad range of measures, each addressed to a bank practice or shortcoming that was believed to have contributed to the financial crisis. These measures were mostly directed at regulated banking organizations—particularly the largest banks, which were perceived as being treated by government as if they were too-big-to-fail. This focus was understandable given the extraordinary steps taken by the government in response to the actual or potential failure of numerous large financial firms in 2008–2009, including brokered mergers and injections of public capital. Dodd–Frank also includes provisions aimed at the derivatives markets. But otherwise, relatively little attention was paid to potential risks generated outside the perimeter of conventional bank holding companies. This emphasis on regulated banking organizations was again understandable, because much of the pre-crisis shadow banking at least indirectly involved institutions that were regulated banks or, in the case of the surviving large investment banks, had turned themselves into regulated banks.

The Dodd–Frank legislation left to regulatory agencies the task of elaborating its often generally stated standards into detailed regulations. Indeed, by the time Dodd–Frank was passed, the US banking regulatory agencies were well along in a self-imposed task of raising capital requirements and imposing liquidity requirements

under existing authority to ensure safety and soundness of banking activities—an enterprise of which Congress was fully aware and appeared to approve.

Accomplishments of Post-Crisis Regulation

While it is obviously too soon to render a complete verdict on the post-crisis regulatory response, three significant accomplishments seem reasonably apparent: tiering of bank regulations by size of institution, greater financial resiliency for bank-related financial institutions, and movement toward an orderly resolution mechanism for failing banks.

Tiering

The 2010 Dodd–Frank legislation established a principle that prudential regulation should vary with the size and systemic importance of banking organizations, based on the magnitude of the negative externalities that would be associated with the stress or failure of various groups of banks. This “tiering” principle was most clearly stated in the Dodd–Frank provision that requires “more stringent” capital, liquidity, risk management, and other standards for banking organizations with more than \$50 billion in assets, explicitly “to prevent or mitigate risks to the financial stability of the United States” (Section 165(a) of the Dodd–Frank Act, 12 US Code §5365). Numerous other requirements apply only to banks above a certain size, such as a requirement that a bank’s financial resilience be evaluated by calculating how it would perform under a variety of “stress testing” scenarios, and the so-called “Volcker Rule” prohibition against a bank engaging in proprietary trading.

Where the thresholds for application or increased stringency of various regulations should be set has been a continuing source of debate since then. The Financial Regulatory Reform Act in May 2018 that raised the threshold for introducing additional prudential measures from \$50 billion in assets all the way to \$250 billion probably went too far, but it also implicitly affirmed that the risks faced and created by banks differ substantially, and that effective and efficient regulation should accordingly vary among groups of banks as well. Going forward, this tiering concept should contribute to a better allocation of both the risk management resources of banks and the supervisory resources of financial regulators.

Greater Financial Resiliency for Bank-Related Financial Institutions

The post-crisis regulatory regime has led to a dramatic increase in the resiliency of the prudentially regulated part of the financial system, as measured by 1) the quality and quantity of capital both required and actually maintained by banks; 2) the greater stability of funding sources for banks; and 3) the risk management capacities and practices of banks. This increase in resiliency extends to the surviving large, formerly “free-standing” investment banks, which are now parts of banking holding companies. Because, *at present*, there is only a moderate amount of runnable short-term funding outside very large banking organizations, the

enhanced resiliency of this group of large banking organizations goes a good way toward containing the risk of a major financial crisis.

Capital requirements were traditionally conceived as helping to combat the moral hazard associated with deposit insurance and lender-of-last-resort assistance. (That is, capital requirements were seen as offsetting the unintended consequences of government programs to staunch bank runs.) Today, they are also recognized as an especially supple prudential tool, insofar as they are available to absorb losses from sources both anticipated and unanticipated by bankers and regulators. Although high capital levels may not in themselves prevent unsustainable increases of credit and associated asset price shocks, they mitigate the severity of the negative externalities associated with those shocks. A bank with high capital will be better positioned to continue providing credit in a crisis. It is also less likely to face a need to sell assets at depressed fire sale prices, which can spread distress to other firms.

Capital requirements may be either a simple ratio of capital to assets or “risk-weighted.” The simple measure, referred to as a “leverage ratio,” has a numerator consisting of a bank’s common equity and some other forms of loss-absorbing capital such as certain preferred stock, and a denominator consisting of all assets.² As the term suggests, a “risk-weighted” capital ratio is calculated by dividing bank capital by a dollar value of assets, other exposures, and off-balance-sheet items that has been adjusted for the perceived riskiness of each asset, determined on the basis of past experience. For the last 30 years, US bank regulations have included both kinds of requirements, on the premise that each compensates in part for the shortcomings of the other. Most regulators here and abroad believe that the risk-weighted requirement should usually be the binding one, while the leverage ratio should help protect against big increases in the riskiness of asset classes above historic norms.

Several important changes have been made to regulatory capital requirements in the post-crisis period. Prior to the crisis, there was only an indirect requirement that banks have minimum amounts of common equity, which can most dependably absorb unexpected losses; some hybrid instruments and even certain forms of subordinated debt qualified as “capital.” During the crisis, though, bank equity was the capital metric of most, and often sole, interest to investors, counterparties, and analysts. A minimum risk-weighted capital requirement of 4½ percent of common equity has accordingly been added, along with a 2½ percent buffer above the minimum. A firm falling into the buffer range of capital requirements must limit its capital distributions, even if it remains above the minimum levels. Leverage ratio requirements have been increased in the United States (and adopted for the first time in many other countries). For the eight US banks designated as being of global systemic importance (Global Systemically Important Banks, or G-SIBs), surcharges for both risk-weighted and leverage requirements have been added.

²As one of the post-crisis reform measures, the Basel Committee on Banking Supervision established a leverage ratio requirement that includes an asset equivalent approximation of derivative and securities exposures, as well as off-balance-sheet items. The US banking agencies have added this requirement for large banks as a “supplemental leverage ratio.”

The so-called “Collins amendment” in Dodd–Frank requires that even the largest banks meet minimum capital ratios based on standardized risk weights that apply to all banks, rather than rely on an internal ratings-based approach to setting capital levels that uses banks’ own models to determine regulatory capital, with its potential for abuse and mistake.

The Federal Reserve has developed a more risk-sensitive capital measure using the annual stress tests required by the Dodd–Frank legislation (Hirtle and Lehnert 2014). Stress tests involve creating unlikely but plausible severe economic scenarios and then using a supervisory model (maintained by the Fed) to estimate the impact of those scenarios on bank assets and earnings. The large banks subject to the stress are required to limit capital distributions so that even under the scenario conditions they would remain above minimum capital requirements. Thus, even if they were to suffer the projected losses, they could remain viable financial intermediaries. The stress test thus substitutes a more forward-looking projection of capital for the fixed 2½ percent buffer in the point-in-time capital requirements. (For other banks not subject to the stress test, the 2½ percent buffer remains.) The simultaneous and comparable testing of all large banking organizations also gives a more complete picture of potential vulnerabilities across the financial system.³

With the caveat that changes in accounting rules and regulatory definitions make comparisons somewhat imprecise, the Federal Reserve Bank of New York (2018) calculates that the aggregate risk-weighted common equity ratio of the largest US banks increased from about 7 percent in the years preceding the financial crisis to about 13 percent as of the end of 2017. At its nadir during the crisis, the risk-weighted capital ratio of this group of banks was barely above 4 percent. The leverage ratio for this same group of banks stood at just under 9 percent at the end of 2017, reversing a downward trend in the decade preceding the onset of the crisis to about 6 percent. Individual ratios for the eight US institutions that have been designated as of global systemic importance are reported in Table 1.

Substantial as this improvement has been, has it been enough? Cline (2017) concludes from both a careful review of the literature and his own cost–benefit analysis that capital levels should be even higher. A study by Federal Reserve Board researchers reaches a similar conclusion (Firestone, Lorenc, and Ranish 2017). The robust stress-testing program in the United States has effectively raised requirements above the nominal, point-in-time minimum ratios and buffers described above, though not as high as the approximately 14 percent recommended by Cline and the Federal Reserve researchers. Indeed, the efficacy of the stress test program itself is still limited by the fact that second-order losses, such as would occur from funding disruptions or a need for forced sales of assets at low fire sale prices, have not yet been incorporated into the supervisory model.

³Among other things, tying capital requirements to stress testing helps counteract, but does not eliminate, the traditional problem of capital ratios as a lagging indicator of bank difficulties. Banks, and sometimes their supervisors, have often postponed the recognition of losses and, thereby, maintained that capital levels were higher than turned out to be the case.

Table 1
**Capital Ratios of the Eight US Firms of Global Systemic Importance
(as of December 31, 2017)**

| <i>Firm</i> | <i>Common equity ratio</i> | <i>Leverage ratio</i> |
|-------------------------|----------------------------|-----------------------|
| Bank of America | 11.9 | 8.6 |
| Bank of New York Mellon | 11.9 | 6.6 |
| Citigroup | 10.3 | 8.8 |
| Goldman Sachs | 12.1 | 8.4 |
| JPMorgan | 12.2 | 8.9 |
| Morgan Stanley | 16.5 | 8.3 |
| State Street | 11.9 | 7.3 |
| Wells Fargo | 12.3 | 9.4 |

Source: Board of Governors of the Federal Reserve, Dodd–Frank Act Stress Test 2018: Supervisory Stress Test Methodology and Results (June 2018).

Notes: The eight banks included in this table are the US firms designated by the Financial Stability Board as of global systemic importance. US banking regulations apply certain requirements only to these eight firms. Under US banking regulations, the Common Equity Ratio is referred to as the “Common Equity Tier 1 Capital Ratio.” Also, under US banking regulations, the Leverage Ratio is referred to as the “Tier 1 Leverage Ratio.”

With respect to the stability of funding sources, a “liquidity-coverage ratio” regulation requires that the largest banking organizations be able to self-fund for 30 days in a period of stress. This is important because it should provide government authorities with a little time in which to consider how to respond to a possibly crippled large bank. This breathing space stands in contrast to the need to devise a plan over a weekend after Bear Stearns and Lehman experienced runs in the spring and fall of 2008, respectively. However, this liquidity-coverage ratio does not address issues of funding sustainability more generally. Through its supervisory oversight, the Federal Reserve supplements this regulatory requirement with annual quantitative liquidity assessments that are customized to the activities and funding needs of individual large banks.

The Federal Reserve estimates that large banks have reduced their reliance on short-term wholesale funding from an amount equivalent to about 50 percent of total assets in the pre-crisis period to about 30 percent today (Quarles 2018a). These banks have roughly tripled their holdings of “high-quality liquid assets,” defined to include reserves at the Fed, Treasuries, and agency securities (Ihrig, Kim, Kumbhat, Vojtech, and Weinbach 2017). As with increased levels of capital, their reduced vulnerability to runs by short-term debtholders has made the financial system more stable. As I will discuss later in this article, however, requirements on funding and liquidity need more work, perhaps considerably more.

The final ingredient for greater resilience of bank financial sheets is that banks have been required to develop and maintain rigorous risk management systems that aggregate information and monitor risk across all business lines. During the first stress test, conducted in early 2009, bank supervisors noted with dismay the serious shortcomings of many banks in quite basic risk management essentials—such as

being able to identify readily the exposures of all a bank's business lines to a specific counterparty. Many banks were especially deficient in projecting and planning for tail risks. As part of the stress-testing process, supervisors required much more attention to these matters and have subsequently observed considerable improvement in these and other risk management capabilities.

An Orderly Resolution Mechanism

In many respects, creating at least the credible possibility of an orderly resolution of a large failing financial institution is the holy grail of efforts to contain the too-big-to-fail problem. If investors and counterparties believe the government would allow even the largest financial firm to fail and the government could in fact do so without endangering the financial system, three beneficial consequences would follow: the moral hazard issue of financial firms taking on excessive risk in the expectation of a government bailout would be substantially contained; crisis amplification effects arising from the prospect of serial failures would be limited; and taxpayer bailouts would be averted.

The 2010 Dodd–Frank legislation created a special insolvency mechanism under which the Federal Deposit Insurance Corporation could manage a large financial firm's resolution in a manner roughly akin to its authority to resolve insured depository institutions, including access to a funding line from the Treasury to inject any needed liquidity into the failed firm. The law also required a planning process for addressing the many practical impediments to resolution, such as organizational complexity, funding shortfalls, cumbersome shared services arrangements, and the prospect that foreign authorities would not permit capital and liquidity to be moved from an American bank's foreign subsidiaries to other subsidiaries in need of resources (so-called “ring-fencing”).⁴ Finally, and in further pursuit of this goal, the Federal Reserve has required the eight systemically important banking organizations to hold significant amounts of longer-term debt that is designated as available for conversion to equity in the event of a bank's failure.

The joint efforts of the Federal Reserve and the Federal Deposit Insurance Corporation (FDIC) in implementing these provisions have yielded some progress in making the “rapid and orderly resolution” of a very large financial firm a credible prospect, though Goldberg and Meehl (2018) observe that the biggest US banks remain quite complex. A natural research approach in looking for evidence on this subject is to examine how cheaply large financial firms are able to raise funds. Were large firms able to raise funds more cheaply than smaller banks before the crisis, based on the expectation that a government bailout would be forthcoming if necessary? Has any such funding advantage diminished in recent years?

⁴ Technically, the resolution planning requirement is that systemically important firms demonstrate that they could be resolved in an orderly fashion *in bankruptcy*—that is, not under the special orderly liquidation authority created by Dodd–Frank. Under existing bankruptcy law, which does not take account of the unusual features of financial firms relative to nonfinancial firms, this aim is an ambitious one. Jackson (2015) and others have proposed amending the US Bankruptcy Code to make it a more viable option for large failing financial firm.

It is notoriously difficult to fix on a point estimate of a pre-crisis funding advantage for very large firms based on market expectations of a government bailout. However, studies of pre-crisis funding conditions such as Ueda and Weder di Mauro (2013) and Brewer and Jagtiani (2013) were consistent in finding a nontrivial positive number. More straightforward in documenting at least the perception of too-big-to-fail was the considerable “uplift” in the pre-crisis debt ratings by major ratings agencies, indicating an expectation of government support. Indeed, the dogged efforts by government authorities during the crisis to avoid outright failure of large financial firms—including government assistance—and the impact on the financial system when Lehman was allowed to fail vindicated the expectations of the market and ratings agencies that bailouts would be forthcoming.

Since the financial crisis, the ratings uplift provided by the credit agencies has dropped significantly. Cetorelli and Traina (2018) find the effective subsidy of the cost of capital of large banks, while still notable, has been reduced (mostly because of an increase in the cost of equity). On the other hand, Afonso, Blank, and Santos (2018) find that while the gap between the credit ratings of bank holding companies and operating subsidiaries has widened, as one would expect if subsidiaries are to be recapitalized using debt at the holding company level, a comparison of bond spreads between parent and subsidiaries does not show a similar narrowing. It appears, then, that the most one can say is that market indicators show a measure of progress, with some indications that investors are pricing in the possibility of failure to a greater extent.

Is an orderly resolution of a huge failed financial institution actually a practical option? My own judgment is that we are probably within shouting distance of the goal of a credible orderly resolution in the case of an idiosyncratic failure of a very large banking organization. In a situation of systemic stress, it is *perhaps* realistic to think we can reach a point at which the first large firm to falter in a period would be placed in resolution, followed by broad measures to inject liquidity into the whole financial system. But the risks of an untested resolution regime are real, and officials may not be willing to take even a modest chance that a systemically important firm placed into resolution would implode. Also, they may be reluctant to use the cache of convertible long-term debt to recapitalize the firm if doing so would impose losses on politically sensitive entities such as pension funds.

Officials might instead look for alternatives, such as an arranged purchase of the failing firm by a stronger bank, quite possibly with government assistance. In 2008, for example, Washington Mutual and Wachovia were both essentially depository institutions for which the Federal Deposit Insurance Corporation already had statutory resolution authority, and neither had significant capital market activities carrying the potential for quick contagion. But instead of placing them into receivership under the FDIC’s authority to resolve depository institutions, the government facilitated their purchase by other, healthier firms (JP Morgan and Wells Fargo, respectively). Finally, it seems very unlikely that multiple large firms would be placed into resolution during a period of high stress. The risks to the financial system would almost surely motivate top government officials to

seek ways to stabilize the system instead. In this sense, too-big-to-fail may remain with us.

Consolidating and Extending the Post-Crisis Framework

The accomplishments of post-crisis financial regulation are substantial, but they have focused mainly on addressing the vulnerabilities and risky practices of banking organizations, and mainly through the exercise of discretion by the banking agencies under quite generally stated statutory requirements or grants of authority. This raises two concerns. First, the increased resiliency of large banking organizations may become degraded over time. Second, less attention has been paid to the risks to financial stability that may arise in the “shadow banking” area—that is, nonbank financial firms that borrow and lend, but do so outside the perimeter of prudentially regulated firms—especially if their borrowing is heavily short-term.

A First Risk: Degradation of the Resiliency of Large Banking Organizations

The future degradation of the resiliency of large banking organizations is by no means foreordained, but it could arise through some combination of the sheer mass of the post-crisis regulatory structure and more-or-less intentional efforts by regulators, in whom so much discretion has been lodged. The 2010 Dodd–Frank legislation called for literally hundreds of new regulations, an approach that entailed protracted and often complicated rulemakings. Many were slowed by the novelty of the new measures, the enormous complexity of measures like the resolution planning process, and the unusual requirement that three, five, or sometimes as many as seven agencies all agree on the regulatory text. While the jury is out on whether this broader participation resulted in better rules, there is no question it stretched out the post-crisis reform efforts considerably. Indeed, eight years after the passage of Dodd–Frank, and nearly a decade after the banking agencies began to work on stronger capital and liquidity regulation, the new regulatory framework in place is still not completed, with numerous proposed regulations not having been finalized.⁵

Dodd–Frank affords substantial discretion to the regulatory agencies. The merit of this approach is that it allows for a more finely tuned and informed regulatory implementation. But it also poses real challenges for the financial regulators, faced with crafting, monitoring, and presumably modifying these regulations as conditions change. For example, core capital and liquidity regulations need continual refinement, both to combat arbitrage efforts by the banks and to adapt to new financial products and conditions. Effective stress tests require constant and timely work

⁵For example, despite the requirement of section 956 of the Dodd–Frank Act that the banking and market regulators adopt regulations to prohibit incentive compensation practices by financial firms that could lead to risky practices, the proposed regulation issued in 2014 has yet to be finalized, and at least four of the relevant agencies removed the proposal from their regulatory agenda in 2017.

to refine loss functions, to guard against incentives for cross-firm asset correlation, and to incorporate scenario elements based on changing economic conditions and financial vulnerabilities.

Moreover, this agency discretion need not always result in more stringent financial rules. Discretion works in both directions, and could allow the rigor of the regulatory system to be substantially reduced without legislation. Rationalization of excessively complicated or unnecessarily burdensome regulation, which nearly everyone agrees is needed to a greater or lesser degree,⁶ could morph into a troublesome deregulation. While Congress legislated some changes to Dodd-Frank in early 2018 that eliminated regulatory requirements for small and mid-sized banks, the banking agencies have moved toward relaxing regulation for the largest banks as well. They have proposed reductions in the leverage ratio surcharge for the eight banks designated as being of global systemic importance (Department of the Treasury, Office of the Comptroller of the Currency, Federal Reserve Board 2018a) and effective reduction in capital requirements for the three “super-regional” banks (Department of the Treasury, Office of the Comptroller of the Currency, Federal Reserve Board 2018b).

The Federal Reserve has also recently proposed integrating point-in-time and stress test capital requirements, along with some changes in stress test assumptions (Board of Governors of the Federal Reserve 2018). As proposed, this step would likely increase modestly the risk-weighted requirements of the largest banks, but it would also effectively reduce leverage ratio requirements. Because the Federal Reserve provided an estimate of the impact of the change only for risk-weighted requirements and did not publicly provide an estimate of the effective decrease in leverage requirements, it is difficult to determine just how much these changes would net out in terms of total capital required for a firm. It remains to be seen whether the Federal Reserve will also respond favorably to the banks’ request for reduction in the amount of the risk-weighted capital surcharges, in which case the amount of capital effectively required by the stress tests would likely decrease.

Even without changes in regulations, stress testing could be made less taxing and less useful. The Federal Reserve’s Vice Chair for Supervision has indicated receptivity to various requests of banks to provide more information to the banks about the supervisory model used to calculate losses and revenues and to find a way to smooth out the impact of the stress tests on banks’ required capital from year to year, because of the variation in the shocks included in the scenario (Quarles

⁶In this context, “rationalization” need not refer only to simplifying certain regulations. It might also be more far-reaching, such as determining whether multiple constraints can achieve similar degrees of protection from financial instability at lower levels (and thus cost to the economy in normal times) than would be needed if reliance was placed on a single metric. For an example of rationalizing through simplification, see the proposal by Greenwood, Hanson, Stein, and Sunderam (2017) to eliminate the leverage ratio while adjusting the stress test approach to risk-weighted assets. For an example of rationalizing through using multiple complementary regulations, based on a retrospective look at how differing capital and liquidity ratios would have performed individually and together pre-crisis, see Aikman, Haldane, Hinterschweiger, and Kapadia (2018).

2018b). Both changes would make capital planning easier for banks, but would also undermine the information and regulatory value of supervisory stress tests. Depending on how these ideas are implemented, considerable reduction in the effective capital requirements of the largest banking organizations could follow. This outcome is made more likely by the vice chair's proposal to eliminate the requirement that large banks' leverage ratio be projected to remain above minimum levels post-stress, since this has been the binding constraint on several of the largest banks in each recent stress test.⁷

Risks of Shadow Banking

“Shadow banking” includes a broad range of nonbank intermediation activity, many forms of which pose little or no threat to financial stability. However, to the extent that post-crisis regulation makes riskier activities more costly for banks, the incentive for these activities to migrate to shadow banking grows. The resulting financial stability concerns will be most tangible where funding is potentially unstable. Of course, a run on nondeposit, “shadow” funding was a central feature of the financial crisis itself, most notably in the repo markets (Gorton and Metrick 2012). Investment banks and others had continually rolled over short-term borrowings to fund longer-term assets. When that funding abruptly dried up as the value of the collateral for that borrowing (such as mortgage-backed securities) was called into question, the effect was similar to a run on bank deposits in a prudential institution.

While this “rollover risk” associated with short-run sources of finance has been only modest in the immediate post-crisis years, there is no guarantee this will remain the case. Yet the post-crisis regulatory approach has created neither a structural solution to shadow banking—for instance, by subjecting all forms of bank-like financial intermediation to a specified regulatory framework—nor the discretionary authority that would enable at least ad hoc responses. There is no generalized authority lodged in the Fed or any other agency to regulate forms of shadow banking that might pose threats to financial stability. It is not clear that there is authority *anywhere* within the US government to regulate the involvement in shadow banking of certain kinds of financial institutions—including hedge funds, private equity funds, and some finance companies.

The Dodd–Frank Act did create a Financial Stability Oversight Council (FSOC), charged with identifying and responding to risks to financial stability. Despite this broad remit, the FSOC's authority is quite limited. Its only direct regulatory powers are the designation of nonbank systemically important financial firms for supervision by the Federal Reserve and the identification of systemically important financial market utilities and payments activities. In its early years, the FSOC designated four

⁷There is continuing debate over whether it is optimal for a capital regulation framework to be designed in such a way that the leverage ratio is the binding constraint under most circumstances. But many who take the position that it generally should not be (myself included) argue that the remedy is to raise risk-weighted requirements, not to lower overall capital in a bank by reducing or eliminating the leverage ratio.

nonbank financial firms as systemically important, and accordingly subject to supervision by the Federal Reserve Board. Subsequently, all have had the designation removed. With political opposition and one federal court's ill-conceived imposition of high barriers to designation, this authority may well be a dead letter for the foreseeable future (Kress 2018).

All other statutory duties of the Financial Stability Oversight Council (FSOC) involve analysis, discussion, reporting, or making recommendations for action to its constituent regulatory agencies.⁸ Because most voting members of an agency that *do* have authority to act on a specific financial stability risk are not represented on the FSOC, a conclusion by the FSOC that a regulatory measure is warranted is only the starting point for what is in effect a negotiation with that agency.

The agency with the existing authority likely to be most salient in addressing shadow banking is the Securities and Exchange Commission. But at least to date, most SEC staff and Commissioners have maintained that the focus of the agency should be its explicit mandates for investor protection and market operations, and that the financial stability mandate for the SEC is quite limited. For example, money market mutual funds are widely considered to be among the ongoing forms of shadow banking that have the potential to produce runs. However, only after prolonged and contentious debate did the SEC agree that institutional funds are prohibited from maintaining a stable net asset value unless they invest only in short-duration government securities.⁹ And there is some skepticism that this measure truly addresses the risk of runs from money market mutual funds. Moreover, many money market funds have apparently met these new requirements by shifting their investments from bank commercial paper to securities issued by the Federal Home Loan Bank System, which in turn lend to banks (Gordon and Gandia 2014; Anadu and Baklanova 2017). Thus, ironically, the “solution” to concerns about runs from money market funds may end up including *de facto* taxpayer support.¹⁰

Macroprudential Policy and Liquidity Regulation

“Macroprudential” policy refers to financial regulation formulated with a view to the health of the financial system as a whole, rather than to the health of individual firms, no matter how large. By definition, macroprudential policy measures should

⁸Because the law made the Secretary of the Treasury the chair of the Council, it was perhaps inevitable that the agenda of the Financial Stability Oversight Council has been significantly weighted toward the sometimes near-term priorities of the presidential administration rather than longer-term financial stability concerns.

⁹The buffer within which the fund may maintain the stable net asset value before it must “break the buck” and reflect a loss creates an incentive for investors to run so as to be insulated from the initial losses to securities held by the fund.

¹⁰The SEC subsequently also adopted a rule requiring certain liquidity risk management practices by asset managers other than money market funds, though the effective date for that rule has now been delayed to late 2018.

cover both bank and nonbank actors; thus, the limits of shadow banking regulation handicap development of macroprudential policies. The case for a macroprudential emphasis had been advanced prior to the financial crisis (Crockett 2000), but was little heeded. As the crisis revealed the consequences of highly correlated asset holdings, shared risks, dependence on runnable short-term funding, and contagion across the financial system, regulators around the world began to invoke macroprudential aims in almost talismanic fashion.

For a number of reasons, macroprudential measures are relatively underdeveloped. While the broad conceptual case for macroprudential measures is strong, substantial analytic work is needed to translate intuitions on system-wide feedback and second-order effects into well-considered and manageable regulatory practice. For example, no real consensus has emerged on the comparative merits of macroprudential policies to increase resiliency of individual institutions versus lean-against-the-wind efforts to prevent unsustainable increases in the price of leveraged assets. Time-varying measures, while conceptually appealing for their countercyclical potential, are especially challenging to specify.

The history of policies that were macroprudential in all but name has not been a particularly happy one in the United States. There is often political opposition to macroprudential policy from legislators and the public, who have regularly pushed back on measures to dampen economic growth in an effort to prevent or avoid relatively rare events (Elliott, Feldberg, and Lehnert 2013). It is perhaps for this reason that, for example, no US government agency has authority to impose a maximum loan-to-value ratio on all mortgages—one of the macroprudential tools most often used in other countries (though its efficacy has been subject to debate, Claessens 2014).¹¹ Institutional concerns are an additional obstacle. As noted earlier, the Financial Stability Oversight Council has no real macroprudential powers, unlike some of its foreign counterparts. As a matter of expertise, the Federal Reserve is probably best equipped to implement macroprudential policies, but there are practical and political reasons not to give it yet more power over the US economy.

Unsurprisingly, then, post-crisis measures with a macroprudential dimension have been directed principally at large regulated banks. The capital surcharges applicable to the eight US banking organizations of global systemic importance are calibrated to take into account the disproportionate impact the failure of one of these firms would have on the financial system as a whole. Along with the other banking agencies, the Fed has put in place a process for imposing additional countercyclical capital requirements in a time-varying fashion, though it has not to date applied such a buffer.

There are also significant macroprudential measures in the annual stress tests. The stress scenarios incorporate a few countercyclical features, such as increasing projected unemployment to a high level even when strong economic growth has

¹¹ Although the banking agencies could apply a rule that limited loan-to-value ratios for regulated banks, the inability to bind nonbank mortgage companies would likely render that action less-than-effective and would surely drive more business to those unregulated firms.

brought actual unemployment to historically low ranges. The supervisory model used in the test measures the effects of stress on the balance sheets of all the larger banks at the same point in time, an approach that comes reasonably close to treating the collective assets of the banking system as a single “portfolio.” Because the Fed requires banks in the stress test to have enough capital to continue lending to credit-worthy borrowers even were the severe scenario to materialize, these features help ensure that recessions would not be intensified as capital-constrained banks stopped lending.¹²

The Financial Regulatory Reform Act of 2018, amending the 2010 Dodd–Frank legislation, may end up reducing the efficacy of this last macroprudential element. While increasing to \$250 billion the dollar asset threshold at which banks would be subject to more stringent prudential requirements, the law kept banks with between \$100 and \$250 billion in assets subject to stress testing. However, it did so through a vaguely stated provision that seems to require less frequent, and perhaps less binding, stress testing. Depending on how the Federal Reserve implements this provision, the roughly \$1.5 trillion in assets held by banks of this size may be removed from the “portfolio” of financial system assets examined annually.

While continued discussion of some form of authority over nonbank lending is worthwhile, the more pressing macroprudential issue is that of liquidity and funding regulation. It is through fragile funding structures that runs begin, and with them the makings of financial crises. A decade after the crisis, the funding profile of large banking organizations looks much healthier. There have been some modest reductions in the vulnerability of the market for repurchase agreements and some changes in the money market fund industry. But the regulatory system has not produced a cohesive set of measures to forestall some future variation on the “run on repo” that was a defining feature of 2008. Problems lie both in the shortcomings of existing liquidity regulation of banks and in the role that nonbank funding needs and practices may play under stressed conditions.

The main existing liquidity regulation, mentioned earlier, is the liquidity-coverage ratio requirement that systemically important banks maintain 30 days of self-funding. This is an important element of an effective crisis management program and, as such, it fulfills a limited macroprudential purpose. But from the very origin of the liquidity-coverage ratio a decade ago, there has been concern that the regulation could cause banks to hoard their liquidity during stress periods. A bank’s sensitivity to market or regulatory agency reactions to its liquidity-coverage ratio declining below minimum levels (or even declining from higher levels) may lead even the soundest banks to sit on their liquidity, rather than use it to reduce liquidity shortages of its customers and markets more generally. Thus, at least in its

¹²The Federal Reserve had other rules in development prior to the 2016 election, including the modeling of some second-order effects for inclusion in the annual stress tests and a proposal to require minimum margins for securities financing transactions not involving Treasuries, regardless of whether the counterparties were regulated banking organizations. The extent to which the Fed will pursue these initiatives during the Trump administration is unclear.

current configuration, the liquidity-coverage ratio may be contributing to a paradoxical situation in which better preparations for orderly resolution could amplify liquidity squeezes at the onset of a period of stress.¹³

The liquidity-coverage ratio is intended only to provide some breathing space for officials confronted with a potential bank failure. Liquidity strains can (and, in the 2007–2008 period, did) extend well beyond 30 days. To address longer-term funding issues, in 2016 the US banking agencies proposed a version of the “net stable funding ratio” (NSFR) developed in the Basel Committee on Banking Supervision. The Basel Committee had already diluted an earlier version of the NSFR because of concerns that its calibration would prevent banks from providing needed liquidity to customers and markets even in fairly normal times. The banking agencies may well dilute it further if they proceed with its eventual adoption. In itself, rendering the NSFR essentially superfluous may not be a bad thing, since it is not especially useful in promoting sustainable funding patterns over the longer term. It creates liquidity requirements based on a comparison of total funding expected to be available over a twelve-month period with expected total funding needs over that same period. Since funding mismatches and consequent shortages arise within much smaller time increments, the NSFR could be unnecessarily restrictive in some respects while still not preventing funding disruptions under stressed conditions.

What is needed, then, is not simply to weaken the net stable funding ratio into insignificance, but to substitute a framework that builds on the actual funding patterns of large banks and that would be an appropriate complement to capital requirements and lender-of-last resort policies by preventing too much reliance on short-term debt. To achieve this goal in an economically sensible fashion, it will be necessary to take the rest of the financial system into account. The demands on a bank’s liquidity, and the availability of funding to it, are substantially dependent on the reactions of central clearing parties, hedge funds, pension funds, insurance firms, money market funds, and other asset managers. To take one example, preliminary work by researchers at both the Federal Reserve Bank of New York (Cetorelli, Duarte, and Eisenbach 2016) and the Bank of England (Baranova, Liu, and Shakir 2017) suggests that corporate bond funds of asset managers may be vulnerable to liquidity squeezes during periods of falling prices. To the extent that these and other forms of nonbank financial intermediaries create funding risks with potentially systemic consequences, it may be more efficient to require them to internalize at least some of the negative externalities they would create for the financial system under stress. Otherwise, liquidity regulation would need to make the prudentially regulated banks *de facto* insurers of liquidity for the nonbank actors.

¹³The resolution plans of the largest banks mandated by the Dodd–Frank Act contain what is in effect a second form of quantitative liquidity requirement. To facilitate resolution in the face of possible restrictions by foreign (or even domestic) regulators on intracorporate liquidity transfers during a crisis, the Fed and the FDIC have obliged the banks to maintain minimum levels of liquidity within certain subsidiaries.

As the preceding discussion suggests, liquidity regulation is still quite under-theorized on a variety of dimensions: its relationship to capital regulation; its interaction with the expected role of the central bank as lender-of-last-resort; and its impact on financial intermediation, including the availability of safe assets. The subject is almost assuredly the major unfinished business of post-crisis reform. It should be a priority for both academic research and policy development.

The Challenge of Crisis Management

In the aftermath of the financial crisis, extraordinary measures taken by the Fed, the Federal Deposit Insurance Commission, and eventually Congress itself helped keep the financial system from freezing up and stabilized financial institutions whether prudentially regulated or not. Meanwhile, very little was done to provide direct assistance to homeowners who found themselves underwater following the precipitous decline in real estate prices. Part of the impetus—both policy and political—behind the 2010 Dodd–Frank legislation was to reduce the chance of any future “bailouts” of financial firms. One approach to this goal was to increase resilience and thus reduce the possibility that systemically important financial firms would fail at all. As noted earlier, another was to create an orderly resolution authority and a requirement for resolution planning so that large distressed financial firms could be wound up, rather than rescued.

But Congress was concerned that the government would still be tempted to offer emergency loans or other assistance to large financial firms, the prospect of which might create unacceptably high moral hazard. Thus, Dodd–Frank took additional steps to rule out such policies. It pared back some longstanding Fed authority to engage in secured lending to nonbanks during “unusual and exigent circumstances.” Similarly, the discretionary authority of the Federal Deposit Insurance to guarantee the liabilities of banks other than insured deposits was made subject to Congressional ratification, which as a practical matter may mean it is not available during a crisis.

Some have drawn the opposite conclusion from the 2008–2009 experience—namely, that the kit of crisis-fighting tools needs to be augmented rather than diminished. Former Treasury Secretary Geithner (2016) has argued that the crisis powers of the Federal Reserve and the Federal Deposit Insurance Corporation should be restored to their pre-Dodd–Frank state. Moreover, he argues that new powers should be created for the Fed to *buy* assets in a crisis (rather than just lend against them), and for the government to inject capital into failing firms.

The unpredictability of future episodes of financial stress, along with the earlier mentioned imprudence of counting too much on a resolution process to deal with problems at systemically important banks, make the current situation worrisome. Still, at present it is hard to see any agreement to add additional crisis-fighting tools. Views are strong. While the vote on the Dodd–Frank legislation as a whole was entirely partisan, a considerable number of Republicans agree that in order to assure an expectation that market discipline will be imposed on failing firms, any

authority to provide extraordinary liquidity and capital during a crisis should be very limited. Thus, in the absence of a credible commitment mechanism to ensure that an enhancement of crisis-fighting tools would not be followed by relaxation of prudential requirements, there does not seem much prospect of support for new tools—even from legislators who might see their merit were the continued robustness of prudential regulation ensured. Here, then, is another example of how the extensive discretion placed in the agencies by the post-crisis regulatory approach can both advance and undercut financial stability aims. In a future crisis, government officials may face the unappealing choice of using only an inadequate set of tools or of taking action that arguably goes beyond the limits set by Congress.

Conclusion

Within the perimeter of prudentially regulated banking organizations, post-crisis financial regulation has made considerable strides, though liquidity regulation needs more work and capital requirements for the biggest banks should probably be somewhat higher. If, during the next few years, the Federal Reserve, Office of the Comptroller of the Currency, and Federal Deposit Insurance Corporation could successfully rationalize the current regulatory framework without weakening the resiliency measures applicable to the most systemically important firms, the broad post-crisis approach to those firms might be relatively durable, albeit with adjustments reflecting industry changes and electoral shifts.

Unfortunately, this outcome is far from assured, and may not even be the most likely. Regulatory agencies over the next few years may undermine the core regulation of the largest banks to the point that proponents of strong regulation will renew their search for more structural measures, or simply blunter measures, in pursuit of financial stability. When control of the Congress and Presidency shifts, as will happen at some point, such ideas will become live policy options. If the largest banks were to be the biggest beneficiaries of developments in financial and payments technology, as seems plausible, the resulting increased concentration in the banking industry could also motivate a more basic change in the post-crisis regulatory regime.

While there is at least a chance for maintaining the progress toward more resiliency for the largest banks, it is considerably harder to conjure up a benign outcome with respect to financial activity that occurs outside the perimeter of banking organizations. Recycled or new forms of shadow banking will almost surely increase over time, whether from existing nonbank financial firms or from new fintech (financial technology) entrants. Some of these will present risks to financial stability. It would be a cruel irony if the mistake of the 1980s and 1990s were repeated, and banking organizations were relieved of core regulations relating to financial stability so as to preserve their franchise value in the face of new competition. It would be crueler still if Congress and federal financial agencies were to wait for another serious financial dislocation before they turned their attention to new risks from new sources.

The argument offered by nonbank financial services firms that they should not be subject to “banking” regulation is often reasonable, but this fair point is not the same as a conclusion that *no* regulation is warranted. Especially with respect to short-term funding or other business models that can produce liquidity squeezes, some system-wide regulation is needed. But with limited legal authority and the apparently exclusive focus of current regulators on deregulatory measures, the business left unfinished by post-crisis reform is likely to remain so for the foreseeable future.

References

- Afonso, Gara Michael Blank, and João Santos.** 2018. “Did the Dodd-Frank Act End ‘Too Big to Fail’?” *Liberty Street Economics*, Federal Reserve Bank of New York, March 5.
- Aikman, David, Andrew G. Haldane, Marc Hinterschweiger, and Sujit Kapadia.** 2018. “Rethinking Financial Stability.” Bank of England Staff Working Paper 712. February 2018. <https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2018/rethinking-financial-stability.pdf>.
- Anadu, Kenechukwu, and Viktoria Baklanova.** 2017. “The Intersection of U.S. Money Market Mutual Fund Reforms, Bank Liquidity Requirements, and the Federal Home Loan Bank System.” Office of Financial Research Working Paper 17-05.
- Baranova, Yuliya, Zijun Liu, and Tamarah Shakir.** 2017. “Dealer Intermediation, Market Liquidity and the Impact of Regulatory Reform.” Bank of England Staff Working Paper 665, July 14, 2017. <https://www.bankofengland.co.uk/working-paper/2017/dealer-intermediation-market-liquidity-and-the-impact-of-regulatory-reform>.
- Board of Governors of the Federal Reserve.** 2018. Notice of Proposed Rulemaking, April 25, 2018, *Federal Register*, Wednesday April 25, 83(80), pp. 18160–88. <https://www.gpo.gov/fdsys/pkg/FR-2018-04-25/pdf/2018-08006.pdf>.
- Brewer, Elijah III, and Julapa Jagtiani.** 2013. “How Much Did Banks Pay to Become Too-Big-To-Fail and to Become Systemically Important?” *Journal of Financial Services Research* 43(1): 1–35.
- Cetorelli, Nicola, Fernando Duarte, and Thomas Eisenbach.** 2016. “Are Asset Managers Vulnerable to Fire Sales?” *Liberty Street Economics* blog, Federal Reserve Bank of New York, February 18.
- Cetorelli, Nicola and James Traina.** 2018. “Resolving ‘Too Big to Fail.’” Federal Reserve Bank of New York Staff Report 859, June 2018. https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr859.pdf.
- Claessens, Stijn.** 2014. “An Overview of Macroprudential Policy Tools.” IMF Working Paper WP/14/214.
- Cline, William R.** 2017. *The Right Balance for Banks: Theory and Evidence on Optimal Capital Requirements*. Washington, DC: Peterson Institute of International Economics.
- Crockett, Andrew.** 2000. “Marrying the Micro and Macro-Prudential Dimensions of Financial Stability.” Remarks before the Eleventh International Conference of Banking Supervisors, Basel, Switzerland, September 20–21, 2000. <https://www.bis.org/review/r000922b.pdf>.
- Department of the Treasury, Office of the Comptroller of the Currency, and Board of Governors of the Federal Reserve.** 2018a. Joint Notice of Proposed Rulemaking, April 18, 2018, *Federal Register* 83(76): 17317–27. <https://www.gpo.gov/fdsys/pkg/FR-2018-04-19/pdf/2018-08066.pdf>.
- Department of the Treasury, Office of the Comptroller of the Currency, Board of Governors of the Federal Reserve, and Federal Deposit Insurance Corporation.** 2018b. Joint Notice of Proposed Rulemaking, October 21, 2018. <https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20181031a1.pdf>.
- Elliott, Douglas J., Greg Feldberg, and Andreas Lehnert.** 2013. “The History of Cyclical Macroprudential Policy in the United States.” Finance and Economics Discussion Paper 2013-29, Federal Reserve Board, Washington, D.C. <https://www.federalreserve.gov/pubs/feds/2013/201329/201329pap.pdf>.

- Federal Reserve Bank of New York.** 2018. "Quarterly Trends for Consolidated U.S. Banking Organizations." https://www.newyorkfed.org/research/banking_trends.html.
- Firestone, Simon, Amy Lorenc, and Ben Ranish.** 2017. "An Empirical Economic Assessment of the Costs and Benefits of Bank Capital in the US." Finance and Economics Discussion Paper 2017-034, Board of Governors of the Federal Reserve, Washington DC. <https://doi.org/10.17016/FEDS.2017.034>.
- Geithner, Timothy F.** 2016. "Are We Safer? The Case for Strengthening the Bagehot Arsenal." Per Jacobsson Foundation Lecture. <http://www.perjacobsson.org/lectures/100816.pdf>.
- Goldberg, Linda S., and April Meehl.** 2018. "Have the Biggest U.S. Banks Become Less Complex?" *Liberty Street Economics* blog, Federal Reserve Bank of New York, May 7.
- Gordon, Jeffrey N., and Christopher M. Gandia.** 2014. "Money Market Funds Run Risk: Will Floating Net Asset Value Fix the Problem?" *Columbia Business Law Review* 2014(2): 314–370.
- Gorton, Gary, and Andrew Metrick.** 2012. "Securitized Banking and the Run on Repo." *Journal of Financial Economics* 104(3): 425–51.
- Greenwood, Robin, Samuel G. Hanson, Jeremy C. Stein, and Adi Sunderam.** 2017. "Strengthening and Streamlining Bank Capital Regulation." *Brookings Papers on Economic Activity*, Fall, pp. 479–544.
- Hirtle, Beverly, and Andreas Lehnert.** 2014. "Supervisory Stress Tests." Federal Reserve Bank of New York Staff Report 696. November.
- Ihrig, Jane, Edward Kim, Ashish Kumbhat, Cindy Vojtech, and Gretchen C. Weinbach.** 2017. "How Have Banks Been Managing the Composition of High-Quality Liquid Assets?" Finance and Economics Discussion Paper 2017-092, Board of Governors of the Federal Reserve, Washington, DC. <https://www.federalreserve.gov/econres/feds/files/2017092pap.pdf>.
- Jackson, Thomas H.** 2015. "Building on Bankruptcy: A Revised Chapter 14 Proposal for the Recapitalization, Reorganization, or Liquidation of Large Financial Institutions." Chap. 2 in *Making Failure Feasible*, edited by Kenneth E. Scott, Thomas H. Jackson, and John B. Taylor. Hoover Institution Press.
- Kress, Jeremy C.** 2018. "The Last SIFI: The Unwise and Illegal Deregulation of Prudential Financial, Inc." November 5. Available at SSRN: <https://ssrn.com/abstract=3278730>.
- Quarles, Randall K.** 2018a. "Liquidity Regulation and the Size of the Fed's Balance Sheet." Remarks at "Currencies, Capital, and Central Bank Balances: A Policy Conference." Hoover Institution Monetary Policy Conference, Stanford University Stanford, California, May 4, 2018. <https://www.federalreserve.gov/newsevents/speech/quarles20180504a.htm>.
- Quarles, Randal K.** 2018b. "A New Chapter in Stress Testing." Remarks at the Brookings Institution, Washington, DC, November 9, 2018. <https://www.federalreserve.gov/newsevents/speech/quarles20181109a.htm>.
- Ueda, Kenichi, and Beatrice Weder di Mauro.** 2013. "Quantifying Structural Subsidy Values for Systemically Important Financial Institutions." *Journal of Banking & Finance* 37(10): 3830–42.

Prone to Fail: The Pre-Crisis Financial System

Darrell Duffie

The financial crisis that began in 2007 was triggered by over-leveraged homeowners and a severe downturn in US housing markets (Mian and Sufi 2014). However, a reasonably well-supervised financial system would have been much more resilient to this and other types of severe shocks. Instead, the core of the financial system became a key channel of propagation and magnification of losses suffered in the housing market (as emphasized in this journal by Gertler and Gilchrist 2018, and discussed in this issue by Aikman, Bridges, Kashyap, and Siegert). Critical financial intermediaries failed, or were bailed out, or dramatically reduced their provision of liquidity and credit to the economy. In the deepest stage of the crisis in September 2008, the failure of Lehman Brothers was accompanied by large, sudden, and widespread increases in the cost of credit to the economy and significant adverse impacts on real aggregate variables (Bernanke 2018).

In short, the core financial system ceased to perform its intended functions for the real economy at a reasonable level of effectiveness. As a result, the impact of the housing-market shock on the rest of the economy was much larger than necessary.

In this essay, I will review the key sources of fragility in the core financial system. The first section focuses on the weakly supervised balance sheets of the largest banks and investment banks. This failure of financial supervision has been widely, if retrospectively, recognized. As one example, Rich Spillenkothen (2010), director of banking supervision and regulation at the Federal Reserve Board from 1991 to 2006, wrote that “prior to the crisis, career supervisors in the regions and at agency headquarters—primarily at the Federal Reserve, Office of the Comptroller of the

■ *Darrell Duffie is Dean Witter Distinguished Professor of Finance, Graduate School of Business, Stanford University, Stanford, California, and a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His email address is duffie@stanford.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at

<https://doi.org/10.1257/jep.33.1.81>

[doi=10.1257/jep.33.1.81](https://doi.org/10.1257/jep.33.1.81)

Currency (OCC), and SEC—failed to adequately identify and prevent the build-up of extreme leverage and risk in the financial system, particularly in large financial institutions.” In a recent University of Chicago poll, US and European economists were asked to gauge the relative importance of twelve factors contributing to the financial crisis. The factor receiving the highest average importance rating in both the European and the American polls was “flawed financial sector regulation and supervision” (IGM Forum 2017).¹

The greatest danger to the functionality of the core of the financial system was posed by five systemically large dealers: Bear Stearns, Lehman Brothers, Merrill Lynch, Goldman Sachs, and Morgan Stanley. These investment banks were exceptionally highly leveraged and dependent on flight-prone sources of short-term liquidity. William Dudley (2009), the President of the Federal Reserve Bank of New York, wrote: “A key vulnerability turned out to be the misplaced assumption that securities dealers and others would be able to obtain very large amounts of short-term funding even in times of stress. ... This short-term funding came mainly from two sources, the tri-party repo system and customer balances in prime brokerage accounts. By relying on these sources of funding, dealers were much more vulnerable to runs than was generally appreciated.” (For more details, see Duffie 2011.)

My emphasis of these topics should not be interpreted as downplaying other sources of systemic risk within the financial system. In particular, other disastrous weaknesses allowed the collapses of AIG, Fannie Mae, and Freddie Mac. But these firms were less critical to the day-to-day functionality of the financial system than the largest commercial banks and the five large investment banks, especially with respect to the continued operation of backbone payments and settlements systems and the provision of liquidity to financial markets.

The middle two sections of this essay focus on the run-prone designs and weak regulation of the markets for securities financing and over-the-counter derivatives, respectively.

Before concluding, I address the undue reliance of regulators on “market discipline.” In the decade before the crisis, US regulators often argued that market discipline would support adequate levels of capital and liquidity at the major banks and investment banks, and that aggressive regulation was unnecessary or counterproductive. But clearly, market discipline did not work. I examine the interplay of too-big-to-fail and the failure of market discipline. Admati and Hellwig (2013) argue that the socially excessive and weakly supervised leverage of the largest financial institutions was essentially subsidized by the government through the presumption by creditors that these firms were too big to fail. Creditors apparently assumed that the biggest banks were too important to be allowed by the government to fail, and

¹I was one of those polled. The other factors listed, in order of assessed average importance among all economists, beginning with the second-most important, were: underestimated risks (financial engineering), mortgages (fraud and bad incentives), funding runs (short-term liabilities), rating agency failures, housing price beliefs, household debt levels, too-big-to-fail beliefs, government subsidies (mortgages, home owning), savings and investment imbalances, loose monetary policy, and fair-value accounting.

thus creditors would not take losses if any of the largest banks or investment banks were to approach insolvency.

Finally, I point to some significant positive strides that have been made since the crisis: improvements in the capitalization of the largest financial institutions, a reduction of unsafe practices and infrastructure in the markets for securities financing and derivatives, and a significantly reduced presumption that the largest financial firms will be bailed out by taxpayer money in the future. But I will also mention some remaining challenges to financial stability that could be addressed with better regulation and market infrastructure.

Regulators Failed to Safeguard Financial Stability

In hindsight, essentially all relevant authorities agree that the largest US financial intermediaries—and especially the five large investment banks Bear Stearns, Lehman Brothers, Merrill Lynch, Goldman Sachs, and Morgan Stanley—were permitted by regulators to have insufficient capital and liquidity, in the years leading up to the crisis, relative to the risks they took. Authoritative voices supporting this view after the crisis include successive chairs of the Federal Reserve Board (Bernanke 2010; Yellen 2015), the Financial Crisis Inquiry Commission (2011), supervisory experts for the Board of Governors of the Fed and the Federal Reserve Bank of New York (Spillenkothen 2010; Gibson and Braunstein 2012; Beim and McCurdy 2009), and country-report examiners at the International Monetary Fund (2010). Oversight of the capital adequacy of the largest investment banks by the Securities and Exchange Commission was particularly lax (Kotz 2010; Schapiro 2010; Inspector General of the Securities and Exchange Commission 2008; Government Accountability Office 2009; Valukas 2010; Bhatia 2011; Gadinis 2012). The insurance company AIG was not effectively supervised by the Office of Thrift Supervision (Polakoff 2009; Finn 2010). The Office of Federal Housing Enterprise Oversight placed few limits on the risks taken by the two giant housing finance intermediaries, Fannie Mae and Freddie Mac (Acharya, Richardson, Van Nieuwerburgh, and White 2011; Stanton 2009). Relative to other regulators, the Federal Reserve had significantly greater supervisory resources and focus on financial stability, yet failed to uncover solvency and liquidity threats that now, with the benefit of hindsight, seem clear.

Yet in the pre-crisis years, there was no apparent urgency to act. I am unable to offer a simple explanation for this failure. Rich Spillenkothen, director of banking supervision and regulation at the Federal Reserve Board from 1991 to 2006, suggested that regulators may have been concerned that actions against large banks would have roiled financial markets. Calomiris and Haber (2014) take a different tack, referring to broad themes of political economy, including the historical US emphasis on a decentralized banking system. In their words, “financial crises occur when banking systems are made vulnerable by construction, as the result of political choices.”

For the specific case of the weak oversight by the Securities and Exchange Commission of the capital and liquidity of the largest investment banks, I am drawn to consider whether the failure to supervise this risk lies with the SEC’s original

mission to protect the customers of financial firms, which crowded out a parallel focus on financial stability (for a related point, see Kohn 2014). As one sign of the emphasis of the SEC on investor protection over financial stability, the Inspector General of the Securities and Exchange Commission (2008, 2009) filed a voluminous 457-page report on the SEC's failure to uncover the Bernie Madoff Ponzi scheme but a mere 27-page report on the SEC's failure to supervise adequately the largest investment banks.

After the crisis, some financial regulators challenged and revised their old approaches. For example, the Fed added substantial resources and focus to its supervision of the largest financial institutions in part through the creation of the Large Institution Supervisory Coordinating Committee in 2010 (Government Accountability Office 2017; Eisenbach, Haughwort, Hirtle, Kovner, Lucca, and Plosse 2015). As another example, a report from the Office of the Comptroller of the Currency (2013) offers a post-crisis review of its supervisory work. By comparison, the reactions of the Securities and Exchange Commission to outside criticisms of its supervision of risk-taking by investment banks—for example, by the Inspector General of the SEC, General Accountability Office (2009), and the Financial Crisis Inquiry Commission (Sirri 2010), and in other public defenses of its pre-crisis supervision (Sirri 2009)—seem narrow and grudging.

An alternative hypothesis for the ineffectiveness of pre-crisis supervision is that it was simply too difficult for regulators to detect the excessive buildup of risk and flight-prone short-run debt and derivatives in the core of the pre-crisis financial system, especially given significant financial innovation and complexity (as argued in Spillenkothen 2010). Some financial intermediaries strategically circumvented leverage restrictions (Acharya and Schnabel 2009; Begley, Purnanandam, and Zheng 2017). Some regulated firms even took steps to hide their true financial conditions, as exemplified by Lehman's infamous Repo 105 practice (discussed in Valukas 2010; Vitan 2013). But while these impediments to supervision are real, regulators should not have been overwhelmed by them. For example, Eisenbach, Lucca, and Townsend (2016) point out that the "existence of economies of scale in bank supervision that are sufficiently strong to outweigh the effect of enhanced supervision for larger banks. This result also suggests that, in terms of realized hour allocations [by supervisory authorities], banks in our sample do not appear to have grown to be 'too large to be supervised.'"

As yet another plausible explanation for the failure of regulators to control the buildup of systemic risk, Gennaioli and Shleifer (2018) propose that investors and policymakers assigned irrationally low probabilities to disaster outcomes, especially with respect to the performance of the housing market. They write: "The Lehman bankruptcy and the fire sales it ignited showed investors and policymakers that the financial system was more vulnerable, fragile, and interconnected than they previously thought. Their lack of appreciation of extreme downside risks was mistaken." Gennaioli and Shleifer "put inaccurate beliefs at the center of the analysis of financial fragility." They note that the second-most important crisis factor according to a poll of leading economists conducted by the IGM Forum (2017), after "flawed financial sector regulation and supervision," is "underestimated risks." An internal

review of pre-crisis supervision conducted at the Federal Reserve Bank of New York by Beim and McCurdy (2009) reached a similar conclusion: “Banks [in their adverse stress scenarios] were not pushed too far out into the tail of the risk distribution or asked to review their plans for dealing with an industry-wide liquidity or credit risk event, or to demonstrate their ability to handle a significant loss of confidence in the industry or loss of funding industry-wide.”

With these various explanations for pre-crisis supervisory failures as a backdrop, I will turn next to how regulation of the main investment banks worked before the financial crisis, and where it fell short. I emphasize two key themes: 1) regulators placed undue reliance on market discipline; and 2) a requirement for reasonable financial stability is that all key financial regulators clearly accept a financial stability mandate (as argued by Kohn 2014; Beim and McCurdy 2009).

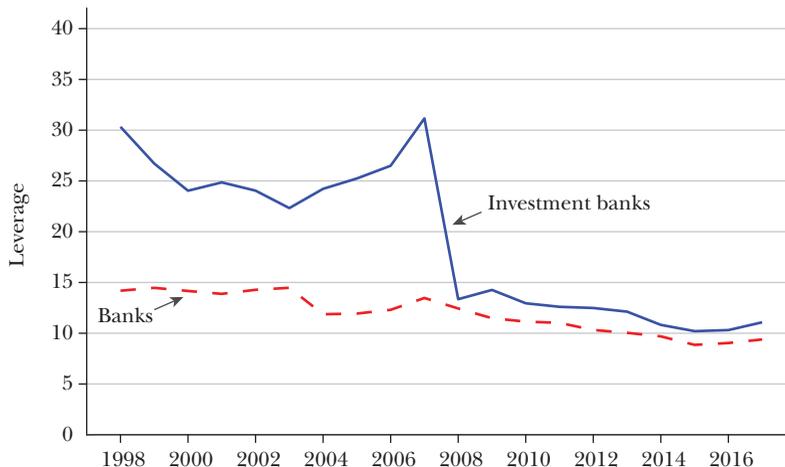
In the years leading up to the financial crisis, the regulatory status of the main investment banks was in some flux. In 2002, the European Union introduced rules that required financial intermediaries operating in the EU to have a consolidated regulatory supervisor. Therefore, all five of these investment banks needed to become supervised by a regulatory agency at the holding-company level. In 2004 and 2005, they elected to be supervised for this purpose by the Securities and Exchange Commission under its new Consolidated Supervised Entity program. In 2008, as the brewing financial crisis came to a full boil, Bear Stearns and Merrill Lynch were forced into mergers with J.P. Morgan and Bank of America, respectively. Lehman Brothers failed. To support their survival, Goldman Sachs and Morgan Stanley became licensed as bank holding companies, giving them direct access to the banking system’s “safety net.” As a result, the SEC shut down its Consolidated Supervised Entity program.

Figure 1 shows the asset-weighted average leverage—that is, the ratio of total accounting assets to accounting equity—of the holding companies of the largest four bank holding companies (J.P. Morgan Chase, Bank of America, Citigroup, and Wells Fargo) and likewise of the five large investment banks (Bear Stearns, Lehman, Merrill Lynch, Goldman Sachs, and Morgan Stanley). The leverage of the investment banks is much higher than shown in the figure at times within each quarter, because they were monitored for compliance only at the end of each quarter (Financial Crisis Inquiry Commission 2011).

The figure clarifies that the SEC’s Consolidated Supervised Entity program was probably not directly responsible for a significant increase in leverage among the investment banks (Sirri 2009). Indeed, Figure 1 shows that the leverage of the investment banks was about as high a decade before the crisis as it was on the opening of the crisis. The SEC’s Associate Director of Trading and Markets, Michael Macchiaroli (2009) emphasized that “the Commission did not relax any requirements at the holding company level because previously there had been no requirements.”²

²A former director of Trading and Markets, Lee Pickard, suggested in 2008 that a 2004 change in the SEC’s minimum net capital rule, Section 15c-3, was responsible for a significant increase in leverage of the investment banks (Securities and Exchange Commission 2004). This assertion is contradicted by Sirri (2009), Lo (2012), and McLean (2012).

Figure 1
Average Leverage (Weighting by Assets) of the Holding Companies of the Largest Investment Banks and Bank Holding Companies



Source: Author using data from SEC 10K filings.

Note: The five investment banks included here are Goldman Sachs, Morgan Stanley, Lehman, Bear Stearns, and Merrill Lynch. The largest bank holding companies are J.P. Morgan Chase, Bank of America, Citigroup, and Wells Fargo. J.P. Morgan Chase merged during the sample period with Bank One and Chase Manhattan. For these calculations, it was treated on a consolidated basis throughout, pro forma, as though these mergers had occurred at the beginning of the sample period.

The extreme leverage of the five investment banks, the existential crises faced by all of them in 2008, and the big post-crisis drop in leverage of the two survivors (Goldman Sachs and Morgan Stanley), all support a view that the SEC had not supervised the investment banks (or their subsidiaries) adequately from the viewpoint of solvency. The Inspector General of the Securities and Exchange Commission (2008) found that the SEC’s Division of Trading and Markets “became aware of numerous potential red flags prior to Bear Stearns’ collapse [in March 2007], regarding its concentration of mortgage securities, high leverage, shortcomings of risk management in mortgage-backed securities and lack of compliance with the spirit of certain Basel II standards, but did not take actions to limit these risk factors.”

As a further illustration of the limited focus of the Securities and Exchange Commission on the solvency of the investment banks, the SEC’s net capital rule (Katz 2004) did not actually constrain the investment banks. The required net capital is 2 percent of “aggregate debt items” (ADI), which is essentially a measure of customer-related claims on the broker dealer subsidiary of the investment bank. There was also an early warning trigger; specifically, the reporting firm is required to notify the SEC whenever the firm’s net capital has breached 5 percent of ADI. Ohlrogge and Giesecke (2016) discuss supplementary forms of capital requirements, but their findings imply that during 2001–2007 the SEC’s net capital requirements represented an average of under 13 percent of the actual net capital reported by

the five investment banks and the broker-dealer subsidiary of Citigroup. Although the investment banks and their subsidiaries had supplementary forms of capital requirements, none of these were effective in controlling solvency risk, nor were they emphasized in SEC supervision.

From a financial stability perspective, a key concern is that the SEC's supervision of risk-taking by the investment banks focused mainly on the protection of the customers of the investment banks from losses rather than on the solvency of their balance sheets and the attendant systemic risks.³ For example, a member of the IMF's country examination staff for the United States wrote that the SEC's mission "stresses ex post enforcement over ex ante prudential guidance" (Bhatia 2011). As another illustration, by my count, only one of a list of 545 pre-crisis SEC regulatory enforcement actions reported in Gadinis (2012) was related to the adequacy of capital or liquidity.⁴ According to the Financial Crisis Inquiry Commission (2011):

Michael Halloran, a senior adviser to SEC Chairman Christopher Cox, told the FCIC the SEC had ample information and authority to require Bear Stearns to decrease leverage and sell mortgage-backed securities, as other financial institutions were doing. Halloran said that as early as the first quarter of 2007, he had asked Erik Sirri, in charge of the SEC's Consolidated Supervised Entities program, about Bear Stearns (and Lehman Brothers), 'Why can't we make them reduce risk?' According to Halloran, Sirri said the SEC's job was not to tell the banks how to run their companies but to protect their customers' assets.

Indeed, the Securities and Exchange Commission devoted few resources to the supervision of the five large investment banks. In September 2008, the SEC's Consolidated Supervised Entity program had a total of only 21 employees supervising these five huge firms, or about four staff members per firm (as noted by Schapiro 2010; see also Financial Crisis Inquiry Commission 2011). By comparison, a very rough estimate based on data from staff reports⁵ of the Federal Reserve Bank

³Ohlrogge and Giesecke (2016) write: "[A] key feature of net capital for broker-dealers is its focus on liquidity, rather than solvency as is the case for bank capital. Calculations of net capital for broker-dealers start with a computation of net worth as defined under generally accepted accounting principles (which thus roughly covers assets minus liabilities, but does not deduct equity). Afterwards, a broker-dealer makes certain adjustments to net worth by adding qualifying subordinated loans, deducting illiquid assets, and then finally applying specified haircuts to the remaining liquid assets in consideration of the market risk they bear. As a result, as the SEC put it, 'net capital essentially means ... net liquid assets.'"

⁴The Gadinis (2012) dataset includes all SEC enforcement actions finalized in 1998, 2005, 2006, and the first four months of 2007 against broker-dealers for any violation of the securities laws.

⁵Table 1 of Eisenbach, Haughwort, Hirtle, Kovner, Lucca, and Plosser (2015) shows that in 2014 the Fed had 22 supervisory staff for each of its "complex financial institutions," which at the time were The Bank of New York Mellon Corporation, Citigroup Inc., The Goldman Sachs Group, Inc., JP Morgan Chase & Co., Morgan Stanley, and the US operations of Barclays PLC, Credit Suisse Group AG, Deutsche Bank AG, and UBS AG, as well as the nonbank firms American International Group, Inc., General Electric Capital Corporation, and MetLife, Inc. From the data underlying Figure 1 of Eisenbach, Lucca, and Townsend (2016), I arrive at a rough estimate of 19 staff per firm in 2008 by multiplying the 2014 number, 22, by the ratio of the total number of full-time equivalent supervisory staff at the Fed in 2008 (which was 583) to the corresponding number in 2014 (which was 671).

of New York is that the Fed devoted about 19 supervisory staff, on average, to each of the systemically important financial firms that it oversaw. The Office of the Comptroller of the Currency also devotes substantial supervisory resources to the largest banks (Office of the Comptroller of the Currency 2013).

The lax supervision of capital adequacy by the Securities and Exchange Commission seemed to be clearly understood by the big investment banks. The Financial Crisis Inquiry Commission (2011) noted,

In January 2008, Fed staff had prepared an internal study to find out why none of the investment banks had chosen the Fed as its consolidated supervisor. The staff interviewed five firms that already were supervised by the Fed and four that had chosen the SEC. According to the report, the biggest reason firms opted not to be supervised by the Fed was the “comprehensiveness” of the Fed’s supervisory approach, “particularly when compared to alternatives such as Office of Thrift Supervision (OTS) or Securities & Exchange Commission (SEC) holding company supervision.

Securities Financing Markets: Core Meltdown Risks

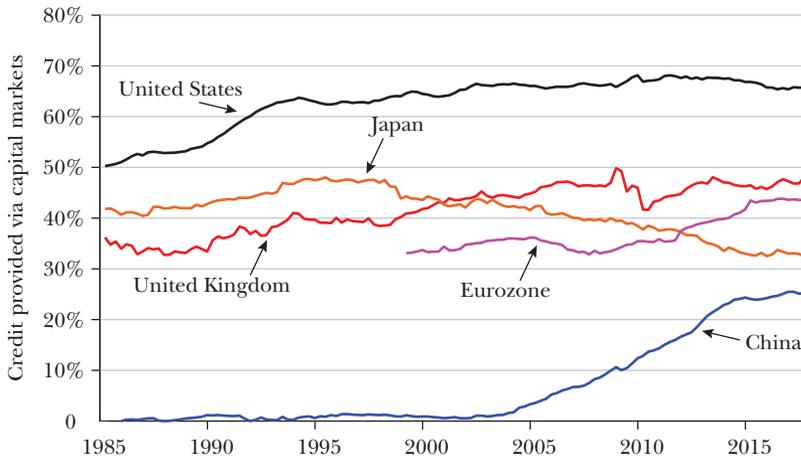
Relative to other major economies and in an absolute sense, credit provision in the United States is significantly more dependent on capital markets than on conventional bank lending. Figure 2 compares the fraction of credit provided via capital markets in several major economies over time. In turn, the intermediation of US capital markets relies heavily on the largest dealers, who make markets by buying securities from investors who want to sell, then selling them to investors who want to buy. Dealers hold securities on their balance sheets in order to provide immediacy to sellers and to have a stock on hand for buyers. Before the crisis, the largest securities dealers (subsidiaries of the investment banks and large commercial banks) financed enormous quantities of inventoried securities with very short-term debt, leaving themselves exposed to risks of creditor runs and fire-sale losses. As famously remarked by Diamond (2013), “[P]rivate financial crises are everywhere and always due to problems of short-term debt.”

The particular crisis of 2007–2009 manifested itself in new forms of short-term debt runs in which repurchase agreements, commonly known as repos, played a major role. Before the crisis, each of the major dealers—again, Goldman Sachs, Morgan Stanley, Lehman, Bear Stearns, and Merrill Lynch—obtained hundreds of billions of dollars in overnight credit in the repo market. On each repo, a dealer transfers securities as collateral to its creditor, and in turn receives cash. When an overnight repo matures the next morning, the dealer is responsible for returning the cash with interest, and is given back its securities collateral.

Money market mutual funds, securities lending firms, and other cash investors in repos often held the collateral securities provided to them by dealers in accounts at two “tri-party” agent banks, J.P. Morgan Chase and Bank of New York Mellon

Figure 2

Fraction of Credit via Capital Markets



Source: Author using data from BIS Statistics Warehouse, at [https://stats.bis.org/#df=BIS:WEBSTATS_TOTAL_CREDIT_DATAFLOW\(2.0\);dq=.CN+GB+JP+US+XM.P.A+B.M+N.XDC.A%3FstartPeriod=1985-01-01&endPeriod=2017-12-01;pv=1,3~7~0,0,0~both](https://stats.bis.org/#df=BIS:WEBSTATS_TOTAL_CREDIT_DATAFLOW(2.0);dq=.CN+GB+JP+US+XM.P.A+B.M+N.XDC.A%3FstartPeriod=1985-01-01&endPeriod=2017-12-01;pv=1,3~7~0,0,0~both).

Note: The fraction shown in the figure is 100 percent minus the percentage of total credit provided by banks to total credit.

(Copeland, Martin, and Walker 2014). Likewise, these repo investors transferred their cash to the dealers’ deposit accounts at the same two tri-party banks.

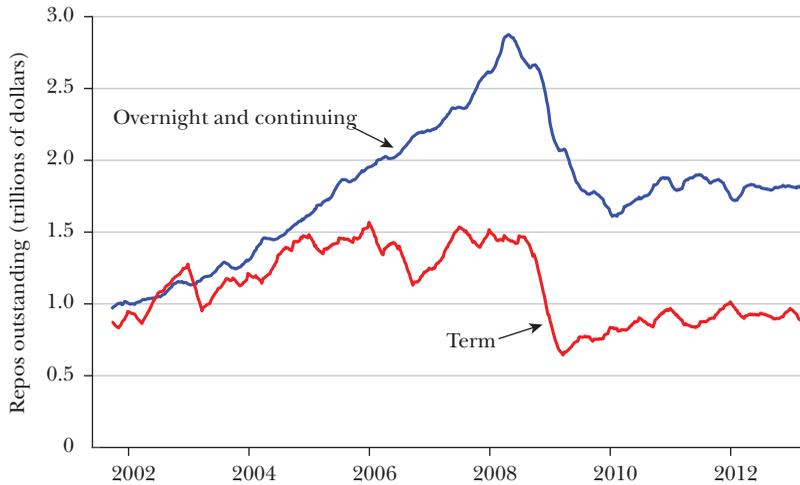
Each morning in the pre-crisis period, when the dealers’ repos matured and they repaid the cash investors, the dealers needed intra-day financing for their securities inventories until new repos could be arranged and settled near the end of the same day. This intra-day credit was provided by the tri-party agent banks. Even “term” repos that had not matured on a given day were temporarily cashed out in the morning and financed during the day by the tri-party banks, a practice that offered operational simplicity. In this manner, up to \$2.8 trillion in intra-day financing was provided to the dealers every day by the two tri-party agent banks (Copeland, Martin, and Walker 2010).

Borrowing in the repo market can either be done on a very short-term basis, such as one day, or on a term basis. Figure 3 shows a significant increase between 2001 and 2008 in the reliance by dealers on one-day repo financing, both in absolute terms and also relative to longer-term repos. This is consistent with the central hypothesis of Gorton, Metrick, and Xie (2014), which is that as financial fragility increased over time, wholesale creditors became more and more anxious to have a quick option to cut their exposures. Of course, this also meant that securities dealers who were continually rolling over their repo agreements, day after day, were vulnerable to the risk that creditors might back away.

This practice is clearly fraught with systemic risk, which is dramatically magnified when key infrastructure providers such as these two tri-party banks are also large sources of credit to their users. This “wrong-way” systemic risk was further

Figure 3

Total Repurchase Agreements Outstanding of US primary dealers
(quarterly rolling averages)



Source: Author using data from Federal Reserve Bank of New York at <https://www.newyorkfed.org/markets/gsds/search.html#>.

Note: “Overnight and continuing” repos are those whose original maturity is one day or which are renewed on a daily basis. Term repos are those with an original maturity of more than one day.

heightened by the practice of settling the cash side of tri-party repos with unsecured commercial bank deposits in the same two tri-party agent banks. These tri-party repo practices exposed the core of the securities funding market to extreme threats in crisis scenarios, and are contrary to well-recognized international standards for financial market infrastructure.⁶ Indeed, since the crisis, an industry task force forced the provision of intra-day credit by the tri-party clearing banks to be almost entirely eliminated (Federal Reserve Bank of New York 2010). However, the practice of settling tri-party repos in unsecured commercial bank deposits persists to this day.

It is useful to spell out how systemic risk can arise in this setting. In the event that a dealer’s solvency or liquidity comes under suspicion, money market funds and other cash investors could decide not to renew the daily financing of the dealer’s securities. This happened to Lehman (Copeland, Martin, and Walker 2014; Krishnamurthy, Nagel, and Orlov 2014).

⁶The settlement of financial market infrastructure transactions in commercial bank deposits is naturally contrary to principles set down by Committee on Payment and Settlement Systems, Technical Committee of the International Organization of Securities Commissions (CPSS-IOSCO) (2012), whose Principle 9 for financial market infrastructure (FMI) states: “An FMI should conduct its money settlements in central bank money where practical and available. If central bank money is not used, an FMI should minimize and strictly control the credit and liquidity risk arising from the use of commercial bank money.” For more details and discussion, see Duffie (2013).

Even if money-fund managers were willing to finance the dealers on a given day, the money fund's own institutional cash investors could run at the first sign of trouble. Moreover, a key SEC regulation governing the composition of money fund assets, Rule 2a7, precludes investment by money funds in the bonds and other assets that they were assigned as repo collateral. Thus, when a dealer fails, its money-fund counterparties could be forced to sell substantial amounts of collateral very quickly—even at fire-sale prices.

If a major dealer was unable to roll over its secured funding during a pre-crisis business day, a tri-party bank's balance sheet would suddenly become imbalanced by the risk of revaluation of hundreds of billions of dollars worth of securities provided by that dealer as intra-day collateral (Duffie 2013). This raised several possible channels for contagion.

First, the tri-party agent banks would have had an incentive (or could be forced by regulations) to sell the collateral securities. A rapid sale would cause a sudden drop in the prices of weaker collateral—of which there was a large amount during the pre-crisis period, including equities and a significant amount of asset-backed securities (Begalle, Martin, McAndrews, and McLaughlin 2016). The spillover of such fire sale prices into security markets and thus onto other investors could have been severe.

Second, under the stress of an intra-day failure by a client dealer, a tri-party agent bank could easily have been prevented from offering tri-party clearing services or intra-day financing to other major dealers. Both operationally and in terms of access to intra-day credit, tri-party repo services were existentially important to the major dealers. With no obvious alternative source of financing, a dealer could have been forced to join the fire sale of securities.

Third, the entire system depended on the willingness of money fund managers and their own sophisticated institutional investors to remain exposed to dealers and to the tri-party repo banks. Institutional investors in “prime” money market funds (those permitted to hold nongovernment securities) are particularly flight prone. As one example, the Reserve Primary Fund disclosed significant losses on investments in commercial paper issued by Lehman Brothers on September 16, 2008. The Fund's net asset value dropped to 97 cents per share, “breaking the buck.”⁷ Within a few days, over \$300 billion of investments in prime money market funds had been redeemed, mainly by “fast” institutional investors (Schmidt, Timmermann, and Wermers 2016; in this journal, see also Kacperczyk and Schnabl 2010). These redemptions occurred even at money funds with little or no exposure to Lehman Brothers. This run on prime money market funds grew in the ensuing

⁷Under post-crisis pressure from the newly created Financial Stability Oversight Council, the Securities and Exchange Commission changed its rules governing money market mutual funds, allowing only those funds investing exclusively in US-government-quality assets to apply “constant net asset value” (CNAV) accounting, which amounts to a fixed price of a dollar a share until rounding forces a fund's net asset value per share below one dollar, thus “breaking the buck.” SEC rules were changed to prevent prime money market funds from using CNAV accounting, and forced these funds to have the ability to apply redemption gates and fees. As a result, over \$700 billion in prime fund investments shifted to government-only money market funds.

days. Absent a halt to this massive flight of one of the main sources of short-term credit to the securities dealers, some or all of these dealers might have been unable to continue financing a substantial fraction of their securities inventories.

In the lead-up to the crisis, an alternative source for substantial amounts of short-term funding was the issuance of “commercial paper” (that is, unsecured debt typically issued for up to six or nine months), either directly or indirectly through off-balance-sheet “structured investment vehicles” (SIVs) created by banks. Baily, Litan, and Johnson (2008) describe the associated liquidity risk as follows.

Until the credit crunch hit in August 2007, this business model worked smoothly: a SIV could typically rollover its short term liabilities automatically. Liquidity risk was not perceived as a problem, as SIVs could consistently obtain cheap and reliable funding, even as they turned to shorter term borrowing ... Technically, the SIVs were separate from the banks, constituting as a ‘clean break’ from a bank’s balance sheet as defined by the Basel II Accord (an international agreement on bank supervision and capital reserve levels), and hence did not add to the banks’ capital or reserve requirements. Once the SIVs ran into financial trouble, however, the banks took them back onto their balance sheets for reputational reasons, to avoid alienating investors and perhaps to avoid law suits.

The asset-backed commercial paper market was particularly prone to runs (Gorton and Metrick 2010, 2012; Gorton, Metrick, and Xie 2014; Schroth, Suarez, and Taylor 2014). A combination of a run on prime money market funds, on other (non-tri-party) sources of repo financing, and on the asset-backed commercial paper market could have caused a complete meltdown of the securities financing market.

Indeed, when such a run began in September 2008, only aggressive action by the Fed and the US Treasury averted an enormous collapse of core financial markets and even deeper panic. The mechanics of this intervention were not straightforward. Securities dealers, including the huge dealer subsidiaries of bank holding companies such as Citibank, Bank of America, and J.P. Morgan, have no direct access to the Fed financing. The Fed’s discount window can provide financing only to regulated banks, and only for “Fed-eligible” collateral, which does not include a significant portion of the assets that were financed in the repo market before the crisis. Moreover, regulatory barriers (Sections 23A and 23B of the Federal Reserve Act) effectively prevent the securities dealer subsidiary of a bank holding company from taking indirect advantage of Fed liquidity that is obtained through the bank subsidiary of the same holding company.

The Fed, lacking other options, invoked its emergency lending authority to provide liberal lender-of-last-resort funding to dealers through a host of new emergency lending facilities: the Term Auction Facility in December 2007; the Primary Dealer Credit Facility in March 2008; the Asset-Backed Commercial Paper Money Market Mutual Fund Liquidity Facility on September 18, 2008; the Commercial Paper Funding Facility on October 7, 2008; the Temporary Liquidity Guarantee Program on October 14, 2008; and the Money Market Investor Funding Facility

on October 21, 2008. Kacperczyk and Schnabl (2010, 2013) offer more details on these programs. The US government and Federal Reserve offered additional crisis support through a vast array of other programs, even to the point of offering a full government guarantee to all money market mutual funds. Without this aggressive fiscal and central-bank support, the impact of the financial crisis on the real economy would have been far deeper than it actually was.

Since the financial crisis, risks associated with the financing of securities by dealers has been reduced in several important ways. I have already mentioned the elimination of intra-day credit provision by tri-party agent banks. The securities inventories themselves are also much smaller, so the need for financing has been correspondingly reduced, as reflected in Figure 3. Because of the declining presumption by bank creditors of “too big to fail,” which I detail later, dealer financing costs have gone up substantially, so the incentive to hold giant inventories is much reduced. The dependence of dealers on flight-prone financing from money market mutual funds has been lowered by a tightening of the regulation of those money funds. Further, bank capital requirements now apply to all large dealers at the holding company level. The two surviving investment banks, which had not been regulated as “banks,” took banking charters and thus became regulated as banks. These capital rules are much more stringent than they were before the crisis, and substantial new bank liquidity coverage regulations have also been introduced, forcing runnable short-term financing to be covered by a stock of high-quality liquid and unencumbered assets.

The Opaque and Unstable Pre-Crisis Swap Market

The enormous pre-crisis over-the-counter derivatives market contributed significantly to the fragility of the financial system. Across the entire over-the-counter derivatives market, there were essentially no regulations governing minimum margin, central clearing, and trade reporting. In practice, the actual amount of margin provided was low (Financial Stability Board 2017). Counterparty exposures and the degree to which they were protected by collateral were generally not observable by anyone other than the two counterparties to each individual position—not even by regulators. Deputy Governor of the Bank of England for Financial Stability Jon Cunliffe (2018) remarked: “The financial crisis exposed complex and opaque webs of bilateral derivatives contracts both between financial firms and with real economy end users. These were often poorly collateralised or not collateralised at all.”

This combination of factors contributed to the risk of a run on major derivatives dealers, which were subsidiaries of the same cast of investment banks and giant commercial banks. In the pre-crisis over-the-counter derivatives market, runs could occur in two main forms. One form was “novation,” a transfer of existing derivatives positions from one counterparty to another. Counterparties of a risky dealer could in some cases use novations to flee to safer dealers. But the most problematic form of run is through the option to terminate over-the-counter derivatives contracts whenever a counterparty experiences an insolvency, a failure to pay, or a change of

control. These run options, legally bypassing bankruptcy rules that force most other types of contracts to stay in place during a reorganization or liquidation, played important roles in the failures of Bear Stearns and Lehman Brothers (Duffie 2011). Derivatives runs drain liquidity and eliminate hedges that are needed by a dealer to manage market-risk exposures. The threat of these runs, as a dealer's position weakens, can cause ordinary creditors to run, a destabilizing feedback that adds to uncertainty over the viability of a dealer, especially given the opaqueness of the dealers' derivatives. In addition to runs, as asset prices related to subprime mortgages fell sharply and concern about counterparty creditworthiness grew, margin calls on derivatives acted as a stress amplifier.

When the largest securities dealers began to fail, their potential exposure to over-the-counter derivatives was huge and opaque, which added to the atmosphere of extreme concern. For example, Cunliffe (2018) notes: "Following its collapse, Lehman's uncleared derivatives counterparties filed claims totalling \$51 billion in relation to its derivatives business. In the event, it was four years before the first payments were made to these uncleared derivatives creditors, and claims against Lehman's are still ongoing." At its failure, Lehman's book of swap positions was actually small in comparison with those of the largest other dealers.

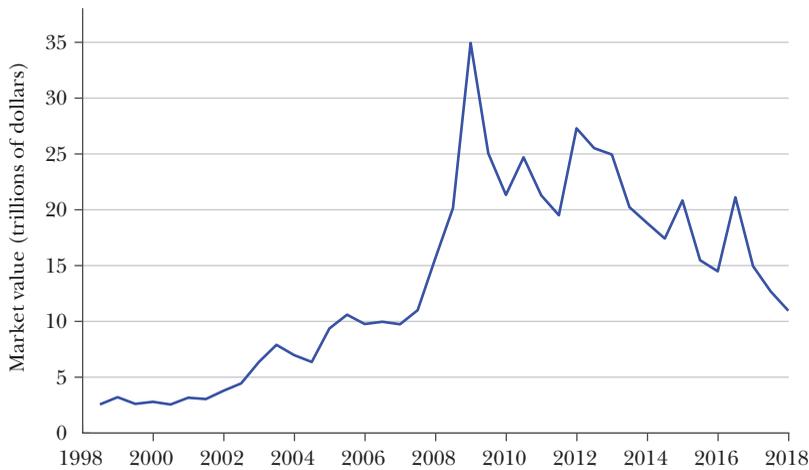
Another form of systemic risk in the derivatives market was caused by AIG's sudden and heavy cash margin calls on credit-default-swap protection that AIG had provided to a number of major dealers on their holdings of subprime mortgages. The dependence of these dealers on AIG's performance on these credit default swaps was an important factor in the decision by the Fed and then the Treasury to rescue AIG (as discussed in this journal by McDonald and Paulsen 2015).

Figure 4 shows a huge pre-crisis buildup in the aggregate gross market value of outstanding over-the-counter derivatives, peaking in 2008 at roughly \$35 trillion dollars. There was ample opportunity before the crisis for regulators to control the buildup of systemic risk in the over-the-counter derivatives market. But when the Commodity Futures Trading Commission (1998) made a move to regulate this market, other regulators pushed back. Treasury Secretary Robert Rubin, Fed Chair Alan Greenspan, and SEC Chairman Arthur Levitt (1998) immediately urged Congress to block the proposed regulation (see also President's Working Group 1999).⁸

Those blocking the regulatory impulses of the Commodity Futures Trading Commission were concerned that new regulations would reduce the legal certainty of over-the-counter derivatives contracts, or would merely encourage a migration of derivatives trading to London. Their concerns led to the passage of deregulatory legislation, the Commodity Futures Modernization Act of 2000, a key step in the striking failure to regulate the enormous build-up of risk in the over-the-counter derivatives market at any time before the crisis (Greenberger 2010). From that point, the size of the over-the-counter derivatives market grew exponentially, and

⁸Rubin, Greenspan, and Levitt (1998) discuss alternative legislation called "Broker-Dealer Lite" under which the Securities and Exchange Commission, and not the Commodity Futures Trading Commission, would regulate the over-the-counter derivatives market.

Figure 4

Global Aggregate Gross Market Values of Over-the-Counter Derivatives

Source: Author using data from Bank for International Settlements (BIS).

Note: For the BIS definition of “gross market value,” see <https://www.bis.org/statistics/glossary.htm?selection=312&scope=Statistics&c=a&base=term>.

with almost no oversight by regulators. One of the “major regulatory and supervisory policy mistakes” identified by Spillenkothen (2010) was the “unwillingness to directly regulate the over-the-counter derivatives market, relying instead on counterparty and market discipline and on supervisors’ assessments of regulated entities’ risk management practices.” McCaffrey (2016) writes:

Many observers view the deregulation of OTC [over-the-counter] derivatives in 2000, through the Commodity Futures Modernization Act, as a serious mistake contributing to the financial crisis. However, no widespread support for external regulation of OTC derivatives existed until after the financial crisis began in 2007. Rather, most analysts accepted on substantive and/or political grounds that the system of private regulation of the OTC derivatives, with informal government oversight, would continue ...

As reflected in Figure 4, post-crisis regulations caused a major decline in the gross outstanding market value of over-the-counter derivatives since the crisis.

A key change is the increased use of central clearing, which was directly mandated in post-crisis regulation and further encouraged by new regulatory capital requirements that, in effect, expressed a preference for central clearing. A central counterparty (CCP), also known as a clearinghouse, enters a derivatives trade as the buyer to the original seller, and as the seller to the original buyer. In this way, original counterparties become insulated from each other’s default risk—provided of course that the clearinghouse meets its own obligations. Central clearing also

improves the transparency of derivatives positions and enforces uniform collateral practices that are more easily supervised by regulators.

An alternative method for lowering counterparty default risk is “compression trading.” By this approach, dealers can eliminate redundant sequences of derivatives positions within the network of dealers that are identified by financial technology companies such as TriOptima. In Duffie (2018), I explain how compression trading has eliminated well over \$1 quadrillion (in notional value) of redundant over-the-counter derivatives. Compression accounts for a substantial portion of the post-crisis reduction in the gross market value of outstanding derivatives shown in Figure 4.

We now know, contrary to concerns expressed in the late 1990s about the potential danger of regulating these markets, that it is possible to add substantial prudential regulation to the over-the-counter derivatives market without stamping out market activity—because this has actually been done in the post-crisis period! Roughly three-quarters of standard swaps are now centrally cleared, all inter-dealer swaps have minimum margin requirements, and all swap transactions must be reported publicly, with details provided to regulatory data repositories that allow the supervision of exposures to individual market participants. Under the Basel-III regulatory capital accord, the largest dealers are now subject to markedly higher capital requirements on their over-the-counter derivatives exposures. Despite these stringent new regulations, potentially useful derivatives trading has not been stifled. In fact, turnover in the over-the-counter derivatives market has continued to rise. For example, the daily turnover for interest-rate derivatives, by far the largest segment of the over-the-counter market, has risen steadily from \$1.7 trillion in 2007 to \$2.7 trillion in 2016 (Bank for International Settlements 2016).

There do remain, however, important concerns over the ability to resolve the failure of central counterparties, which have become enormous concentrations of risk under post-crisis regulations. If a clearinghouse has insufficient resources to manage the default of the derivatives obligations of a clearing member, the consequences could be catastrophic, now that hundreds of trillions of derivatives have been cleared by a small number of systemically important central counterparties. The default management resources of the central counterparty consist primarily of the margins provided by clearing members against their positions, and by a default fund to which all clearing members contribute. If the initial margin of a failed clearing member is not enough to cover the losses, the default fund is then applied. If the clearinghouse burns through both of these paid-in default management resources, and a small layer of its own capital, it then has the contractual right to stop paying clearing members the amounts otherwise due on their derivatives, even to the point of “tearing up” their derivatives positions. In the worst scenarios, the cessation of payments to clearing members and tear-ups would be catastrophic, and contagious. The largest clearing members are generally also large members of other central counterparties. This tail contagion risk is subject to regulatory stress tests and ultimately to regulations that could trigger a failure resolution process for central counterparties. However, actual implementable plans for the failure resolution of clearinghouses have still not been designed, at least in the United States (Duffie 2013, 2015, 2018). Cunliffe (2018) provides an update of regulatory progress in this area.

Too-Big-to-Fail Eviscerates Market Discipline

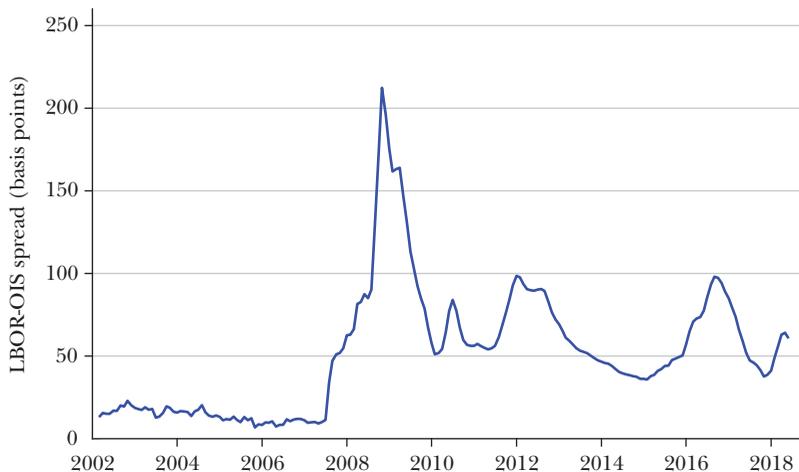
In the decade or so before the financial crisis arrived in 2007, it was common to see claims that market discipline could lead to less government regulation. In 1997, Fed Chair Alan Greenspan stated: “As we move into a new century, the market-stabilizing private regulatory forces should gradually displace many cumbersome, increasingly ineffective government structures. This is a likely outcome since governments, by their nature, cannot adjust sufficiently quickly to a changing environment, which too often veers in unforeseen directions.” In 2000, Fed Governor Laurence Meyer stated: “As large banking institutions become increasingly complex—and fund themselves more from non-insured sources—market discipline and its prerequisite, public disclosure, must play a greater role. Indeed, increased transparency and market discipline can also help substantially to address concerns about increased systemic risk associated with ever-larger institutions and to avoid the potentially greater moral hazard associated with more-intrusive supervision and regulation.” The sentiment was international. For example, the Basel Committee on Banking Supervision (1999) wrote that market discipline “imposes strong incentives on banks to conduct their business in a safe, sound and efficient manner.”

Evidence from the crisis of 2007–2009, however, soundly rejects the power of market discipline to maintain financial stability. As Fed Chair Janet Yellen (2015) acknowledged: “The checks and balances that were widely expected to prevent excessive risk-taking by large financial firms—regulatory oversight and market discipline—did not do so.”

In a post-crisis Congressional hearing, Henry Waxman (D-CA) asked Greenspan, “Well, where did you make a mistake then?” Greenspan replied, “I made a mistake in presuming that the self-interest of organizations, specifically banks and others, were such that they were best capable of protecting their own shareholders and their equity in the firms” (House of Representatives, Committee on Oversight and Government Reform 2008, p. 33). In his prepared remarks, Greenspan (p. 17) similarly commented: “Those of us who have looked to the self-interest of lending institutions to protect shareholders’ equity, myself included, are in a state of shocked disbelief. Such counterparty surveillance is a central pillar of our financial markets state of balance.”

An internal Federal Reserve Bank of New York review of pre-crisis supervisory weaknesses conducted by Beim and McCurdy (2009) offers similar and more pointed criticisms. They describe two “basic assumptions [that] are wrong: 1. ‘Banks can be relied upon to provide rigorous risk control.’ In reality banks’ internal risk management and control functions were often ineffective in the run-up to the crisis and were usually trumped by the pressure to do profitable business. 2. ‘Markets will always self-correct.’ A deference to the self-correcting property of markets inhibited supervisors from imposing prescriptive views on banks.” They wrote: “Interviewees noted the common expectation that market forces would efficiently price risks and prompt banks to control exposures in a more effective way than regulators.”

Figure 5

Average One-Year Credit Spread of Large Banks Borrowing US Dollars: LIBOR versus the OIS Swap Rate

Source: Author using data from Bloomberg.

Note: The figure shows the difference between the one-year US Dollar London Interbank Offered Rate (LIBOR) and the one-year overnight index swap (OIS) rate based on the Fed Funds rate.

Reliance on market discipline implies an assumption that excessive risk-taking by a financial intermediary will be limited by the intermediary's cost of debt financing, based in turn on creditors' perceived risk of losses at insolvency. However, before the financial crisis, there was nothing close to a realistic plan for how to resolve the insolvency of systemically important financial firms without triggering or deepening a crisis. This created a presumption among creditors that the largest banks were "too big to fail."

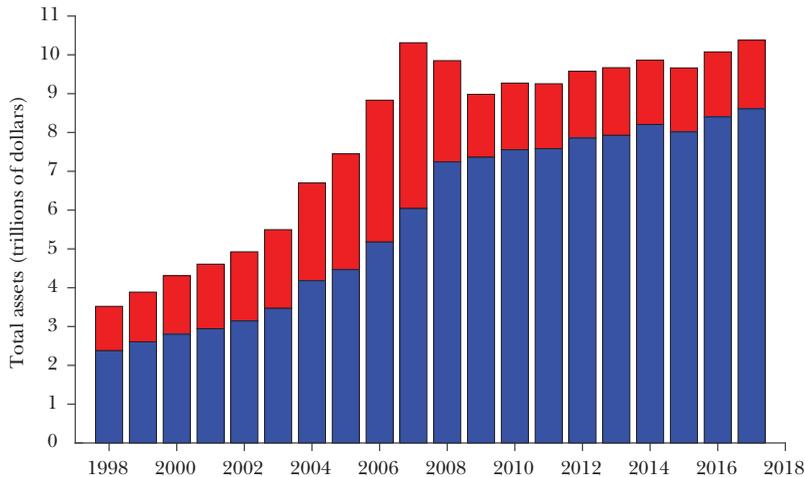
Thus, despite their thin pre-crisis solvency buffers, the big banks and investment banks experienced what is in retrospect an amazingly low cost of credit. As one example, Figure 5 shows the one-year credit spreads of large banks. Here, "LIBOR" (the London Interbank Offered Rate) is the rate at which the largest banks can borrow from each other, while the OIS (overnight indexed swaps) rate is a proxy for the rate of interest of borrowers that are nearly risk-free. The razor-thin LIBOR-OIS credit spread that large banks paid from 2002–2007 shows that their creditors had very little concern about lending to them, right up until the financial crisis hit.

With this low cost of borrowing, the pre-crisis cost to big-bank shareholders of expanding their balance sheets with debt financing was much lower than the associated social costs stemming from systemic failure risk. Their trading desks jumped at almost any opportunity to borrow that allowed them to grab a few basis points of profit, because their funding costs were so low.⁹ Indeed, Figure 6 shows a tripling of

⁹As an example, in Andersen, Duffie, and Song (forthcoming), we model how pre-crisis banks could exploit their exceptionally low credit spreads to capture shareholder profits from even small violations

Figure 6

Total Assets, by Year, of the Five Largest Investment Banks (Red) and the Four Largest Banks (Blue)



Source: Author using data from SEC 10K filings.

Note: The five largest investment banks (the red section of the bars, at the top) are Goldman Sachs, Morgan Stanley, Lehman, Bear Stearns, Merrill Lynch. The four largest banks (the blue section of the bars, at the bottom) are J.P. Morgan Chase, Bank of America, Citigroup, and Wells Fargo. J.P. Morgan Chase merged during the sample period with Bank One and Chase Manhattan. For these calculations, it was treated on a consolidated basis throughout, pro forma, as though these mergers had occurred at the beginning of the sample period.

the total assets of the five largest investment banks and the four largest banks during the decade leading up to the crisis. The incentive to borrow caused by being too big to fail and the lack of methods for safely resolving an insolvency of any of these firms, combined with the forbearance of regulators, created an increasingly toxic brew of systemic risk.

Was the dramatic expansion of borrowing in the financial sector because of moral hazard—that is, an assumption by firms that they would be bailed out? Genaioli and Shleifer (2018) argue against conventional moral-hazard explanations of the excessive pre-crisis leverage of the big banks, and I agree. Instead, the moral hazard explanation applies to creditors, who were apparently convinced that these firms would not be allowed to fail. In expanding their balance sheets with debt, financial firms did not even need to think about the moral hazard of government bailouts—they merely needed to observe the exceptionally low costs of debt financing offered to them by creditors. When Lehman ultimately did fail, the surprise of creditors exacerbated the ensuing panic (Bernanke 2018; Gennaioli and Shleifer 2018).

of covered interest parity (CIP). In the post-crisis era, however, much larger CIP violations remain unexploited because of substantially higher big-bank debt funding spreads.

A primary factor in the judgment of creditors that the largest financial intermediaries were too big to fail was that there was no method for resolving an insolvency that didn't also crater the economy. In a standard bankruptcy procedure, contracts and claims are frozen in place during a liquidation or reorganization. However, the huge books of over-the-counter derivatives and repos of the largest banks and investment banks are legally exempt from the bankruptcy code as "qualified financial contracts" (QFCs). Because of this bankruptcy exemption for QFCs, counterparties to failing financial firms in these contracts are not required by an "automatic stay" to freeze their positions in place. Instead, counterparties can quickly terminate their contracts and keep their collateral (for details, see Duffie and Skeel 2012).

In order for market discipline to limit failure risk, creditors need to believe that they could be forced to experience a significant loss at insolvency. In the future, regulators are planning to use post-crisis legislation—Title II of the US Dodd-Frank Act and the European Union's Bank Recovery and Resolution Directive—to force wholesale "loss-absorbing" creditors to give up their debt claims when a large bank nears insolvency. In effect, these creditor claims are to be cancelled and replaced with equity claims. The threat of invoking this resolution scheme, called "bail-in," is made more credible through new legislation that includes a temporary stay on the termination of over-the-counter derivatives and repos.

Other efforts are being made to improve failure resolution methods (for an update, see US Department of the Treasury 2018). As one example, Jackson (2015) has proposed amending the US bankruptcy code with a new Chapter 14, which is designed to address the failure of systemically important financial institutions. Like Title II of the Dodd Frank Act, Chapter 14 would impose a temporary stay on over-the-counter derivatives and repos.

Whether or not bail-in actually works reasonably well in practice, what matters for big-bank borrowing costs is that *creditors believe that it would be tried*. It appears that they do now believe this. As shown earlier in Figure 5, the cost of wholesale unsecured credit for the largest banks as measured by the LIBOR-OIS credit spreads has increased dramatically and now fluctuates more notably with credit-related events.

Sarin and Summers (2016) argue that higher post-crisis big-bank credit spreads reflect a continuing failure of these firms to improve their solvency. In their view, these high post-crisis credit spreads reflect the reduced franchise values of their business operating models, rather than a reduced reliance by creditors on too-big-to-fail. However, Rosengren (2013), Carney (2014), and Tucker (2014) estimate a full order of magnitude increase in the capital buffers of the largest banks. Similarly, in Berndt, Duffie, and Zhu (2018), my coauthors and I estimate a major improvement in the "solvency ratios" of most large financial firms, defined as the ratio of tangible common equity to an estimate of the standard deviation of the annual change in the market value of the firm's assets. We find that the solvency ratios of the largest financial firms averaged only about 0.3 from 2002–2008, but have risen to around 0.8–1.0 since 2013. We argue that the general post-crisis increase in credit spreads of large financial firms does not reflect a continuing low level of solvency, but instead is a reaction by creditors to the increased probability

that the government would force wholesale creditors of a large bank approaching insolvency to take a significant loss.

A belief by creditors that the largest banks are no longer too big to fail leads to a better alignment of the risk-taking incentives of these banks with social incentives to control systemic risk. The greater is the credit spread of a financial intermediary, the greater is the impact of debt overhang in reducing the incentives of its shareholders to expand the intermediary's balance sheet using debt financing. Indeed, since the crisis, significant increases in unsecured dealer credit spreads have forced the trading desks of the largest dealers to charge their trading clients for newly designated "funding value adjustments." In Andersen, Duffie, and Song (forthcoming), we explain these funding value adjustments as debt-overhang costs to bank shareholders for enlarging their balance sheets. Thus, because of new failure resolution rules, market discipline has to some extent finally begun to work.

Although the incentives of big-bank shareholders to expand their balance sheets are now more aligned with social incentives, day-to-day market liquidity has in some cases suffered, a different form of social cost.

Final Remarks

Leading up the crisis, the core of the financial system was not prepared to withstand a significant shock. An undue reliance on market discipline had left the largest financial firms undercapitalized, and this was exacerbated by a failure of the Securities and Exchange Commission to prioritize financial stability. Core financial firms were actually encouraged, through artificially low costs of debt financing, to use leverage to grow enormous balance sheets. Creditors competed to supply these firms with funding at razor-thin credit spreads because they did not believe that these firms would be allowed by the government to fail. Their belief in "too big to fail" was based on the presumption of large spillover costs of failure on the broader economy. With hindsight, this presumption was correct. When Lehman actually did fail, it was impossible to avoid enormous bankruptcy costs and contagion because safe insolvency resolution methods for large banks had not been developed.

Since the crisis, major strides toward financial stability have been achieved. The largest US dealer banks are all now under the supervision of the Federal Reserve. Their capitalization and liquidity has been forced up with stringent new banking regulations. Some weaknesses in market infrastructure and unsafe practices in the markets for securities financing and derivatives have been corrected. New failure resolution methods now prevent derivatives and other critical financial contracts from suddenly terminating at insolvency. As a result, general creditors to these firms no longer presume that they will be bailed out. This has led to much higher costs of debt financing for these firms, which has discouraged their leverage and has knocked down the rapid pre-crisis growth of their balance sheets.

Challenges to the resilience of the core financial system remain. We do not yet know how well failure resolution methods for the largest banks will actually work in practice. There is still no known operational planning for US government failure

resolution of derivatives clearinghouses. Meanwhile, regulations have forced the majority of derivatives risk into these clearinghouses, which are the new “too big to fail” financial firms. And there will always be a threat that, with the passage of time, fading memories of the costs of the last crisis will lower the resolve and vigilance of legislatures and financial regulators to monitor changes in practice and to take steps to control socially excessive risk-taking.

■ *An earlier version of this paper was presented at “The Financial Crisis at 10,” a symposium of the 2018 Summer Institute of the National Bureau of Economic Research. Because I was on the board of directors of Moody’s Corporation from October 2008 to April 2018, I avoid a discussion of credit rating agencies. I am grateful for research assistance from Marco Lorenzon, Yang Song, and David Yang, and for conversations with and comments from Thomas Eisenbach, Gary Gorton, Joe Grundfest, Anil Kashyap, Michael Ohlrogge, Hyun Shin, Andrei Shleifer, Jeremy Stein, Larry Summers, and Paul Tucker. I am especially grateful for detailed comments and suggestions from Mark Gertler, Gordon Hanson, and Timothy Taylor.*

References

- Acharya, Viral, Matthew Richardson, Stijn Van Nieuwerburgh, and Lawrence J. White. 2011. *Guaranteed to Fail: Fannie Mae, Freddie Mac, and the Debacle of Mortgage Finance*. Princeton University Press.
- Acharya, Viral V., and Philipp Schnabl. 2009. “How Banks Played the Leverage ‘Game’?” Chap. 2 in *Restoring Financial Stability: How to Repair a Failed System*, edited by Viral V. Acharya and Matthew Richardson. New York: Wiley.
- Admati, Anat, and Martin Hellwig. 2013. *The Bankers’ New Clothes: What’s Wrong with Banking and What to Do About It*. Princeton University Press.
- Aikman, David, Jonathan Bridges, Anil Kashyap, Anil, and Caspar Siegert. 2018. “Would Macroprudential Regulation Have Prevented the Last Crisis?” Bank of England Working Paper 747, June.
- Andersen, Leif, Darrel Duffie, and Yang Song. Forthcoming. “Funding Value Adjustments.” *Journal of Finance*.
- Baily, Martin Neil, Robert E. Litan, and Matthew S. Johnson. 2008. “The Origins of the Financial Crisis.” Paper 3, Fixing Finance Series, Brookings Institution. November. https://www.brookings.edu/wp-content/uploads/2016/06/11_origins_crisis_baily_litan.pdf.
- Bank for International Settlements. 2016. “Triennial Central Bank Survey: OTC Interest Rate Derivatives Turnover in April 2016.” Monetary and Economic Department. September. (Annex tables revised, December 2016). <https://www.bis.org/publ/rpfx16ir.pdf>.
- Basel Committee on Banking Supervision. 1999. “A New Capital Adequacy Framework,” Consultative paper issued by the Basel Committee on Banking Supervision, June. <https://www.bis.org/publ/bcbs50.pdf>.
- Begalle, Brian J., Antoine Martin, James McAndrews, and Susan McLaughlin. 2016. “The Risk of Fire Sales in The Tri-Party Repo Market.” *Contemporary Economic Policy* 34(3): 513–30.
- Begley, Taylor A., Amiyatosh Purnanandam, and Kuncheng Zheng. 2017. “The Strategic Underreporting of Bank Risk.” *Review of Financial Studies* 30(10): 3376–3415.
- Beim, David, and Christopher McCurdy. 2009. “Federal Reserve Bank of New York Report on Systemic Risk and Bank Supervision” Draft of August 5, 2009. https://fcic-static.law.stanford.edu/cdn_media/fcic-docs/2009-08-05%20FRBNY%20Report%20on%20Systemic%20Risk%20and%20Supervision%20Draft.pdf.
- Bernanke, Ben S. 2010. “Causes of the Recent Financial and Economic Crisis.” Testimony Before the Financial Crisis Inquiry Commission, Washington, DC. September 2. Board of Governors of

the Federal Reserve. <https://www.federalreserve.gov/newsevents/testimony/bernanke20100902a.htm>.

Bernanke, Ben. 2018. "The Real Effects of Disrupted Credit." Presentation at the Nobel Symposium on Money and Banking, Stockholm, Brookings Institution, Washington D.C. September 13. https://www.brookings.edu/wp-content/uploads/2018/09/BERNANKE_PRESENTATION_BPEA.pdf.

Berndt, Antje, Darrell Duffie, and Yichao Zhu. 2018. "The Decline of Too Big to Fail." December. <https://www.darrellduffie.com/uploads/pubs/BerndtDuffieZhuTBTF.pdf>.

Bhatia, Ashok Vir. 2011. "Consolidated Regulation and Supervision in the United States." International Monetary Fund Working Paper 11/23, January. <https://www.imf.org/external/pubs/ft/wp/2011/wp1123.pdf>.

Calomiris, Charles W., and Stephen H. Haber. 2014. *Fragile by Design: The Political Origins of Banking Crises and Scarce Credit*. Princeton University Press.

Carney, Mark. 2014. "The Future of Financial Reform." 2014 Monetary Authority of Singapore Lecture, Monetary Authority of Singapore, Singapore, November 17. <https://www.bankofengland.co.uk/speech/2014/the-future-of-financial-reform>.

Commodity Futures Trading Commission. 1998. "Over-the-Counter Derivatives." Concept Release, Washington, DC, May. https://www.cftc.gov/sites/default/files/opa/press98/opamntn.htm#p60_8020.

Committee on Payment and Settlement Systems, Technical Committee of the International Organization of Securities Commissions (CPSS-IOSCO). 2012. "Principles for Financial Market Infrastructures." Bank for International Settlements and IOSCO. April. <http://www.bis.org/publ/cpss101a.pdf>.

Copeland, Adam M., Antoine Martin, and Michael Walker. 2010. "The Tri-Party Repo Market before the 2010 Reforms." Federal Reserve Bank of New York Staff Report Number 477, December.

Copeland, Adam, Antoine Martin, and Michael Walker. 2014. "Repo Runs: Evidence From the Tri-Party Repo Market." *Journal of Finance* 69(6): 2343–80.

Cunliffe, Jon. 2018. "Central Clearing and Resolution—Learning Some of the Lessons of Lehman's." Speech at Futures Industry Association (FIA) annual International Derivatives Expo, London. Bank of England. Published June 5. <https://www.bankofengland.co.uk/speech/2018/jon-cunliffe-speech-at-futures-industry-association-international-derivatives-expo>.

Diamond, Douglas W. 2018. "Liquidity Creation by Financial Intermediaries." Presentation,

Nobel Symposium on Money and Banking, Stockholm, May 26. <https://www.houseoffinance.se/wp-content/uploads/2018/03/Diamond2.pdf>.

Dudley, William. 2009. "More Lessons from the Crisis." Remarks at the Center for Economic Policy Studies (CEPS) Symposium, Princeton, New Jersey, Federal Reserve Bank of New York, April. <https://www.newyorkfed.org/newsevents/speeches/2009/dud091113.html>.

Duffie, Darrell. 2011. *How Big Banks Fail and What to Do About It*. Princeton University Press.

Duffie, Darrell. 2013. "Replumbing Our Financial System: Uneven Progress." *International Journal of Central Banking* 9(January): 251–80.

Duffie, Darrell. 2015. "Resolution of Failing Central Counterparties." Chap. 4 in *Making Failure Feasible: How Bankruptcy Reform Can End "Too Big to Fail,"* edited by Kenneth E. Scott, Thomas H. Jackson, and John B. Taylor. Stanford: Hoover Institution Press.

Duffie, Darrell. 2018. "Financial Regulatory Reform After the Crisis: An Assessment." *Management Science* 64(10): 4835–4857.

Duffie, Darrell, and David Skeel. 2012. "A Dialogue on the Costs and Benefits of Automatic Stays for Derivatives and Repurchase Agreements." Chap. 5 in *Bankruptcy Not Bailout: A Special Chapter 14*, edited by Kenneth E. Scott and John B. Taylor. Hoover Institution Press.

Eisenbach, Thomas, Andrew Haughwort, Beverly Hirtle, Anna Kovner, David Lucca, and Matthew Plosser. 2015. "Supervising Large, Complex Financial Institutions: What do Supervisors Do?" Federal Reserve Bank of New York Staff Report 729, May.

Eisenbach, Thomas M., David O. Lucca, and Robert M. Townsend. 2016. "The Economics of Bank Supervision." Federal Reserve Bank of New York Staff Report 769, March.

Federal Reserve Bank of New York. 2010. "Tri-Party Repo Infrastructure Reform." May 17. https://www.newyorkfed.org/medialibrary/media/banking/nyfrb_triparty_whitepaper.pdf.

Financial Crisis Inquiry Commission. 2011. *The Financial Crisis Inquiry Report: Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States*. Official Government Edition, January US Government Printing Office. January.

Financial Stability Board. 2017. "Review of OTC Derivatives Market Reforms: Effectiveness and Broader Effects of the Reforms." Financial Stability Board, Basel, June 29. <http://www.fsb.org/wp-content/uploads/P290617-1.pdf>.

Finn, Michael. 2010. "Statement of Michael E. Finn, Northeast Region Director, Office of Thrift Supervision, regarding American International Group, before the Congressional Oversight Panel." May 26. Office of Thrift Supervision,

Department of the Treasury, Washington DC. <https://occ.gov/static/ots/testimony/ots-testimony-ts189-05-26-2010.pdf>.

Gadinis, Stavros. 2012. "The SEC and the Financial Industry: Evidence from Enforcement Against Broker-Dealers." *Business Lawyer* 67(3): 679–728.

Gennaioli, Nicola, and Andrei Shleifer. 2018. *A Crisis of Beliefs: Investor Psychology and Financial Fragility*. Princeton University Press.

Gertler, Mark, and Simon Gilchrist. 2018. "What Happened: Financial Factors in the Great Recession." *Journal of Economic Perspectives* 32(3): 3–30.

Gibson, Michael S., and Sandra F. Braunstein. 2012. "Consolidated Supervision Framework for Large Financial Institutions." Letters SR 12-17 and CA 12-14, Board of Governors of the Federal Reserve System, Washington, DC. December 17. <https://www.federalreserve.gov/supervisionreg/srletters/sr1217.htm>.

Gorton, Gary, and Andrew Metrick. 2010. "Regulating the Shadow Banking System." *Brookings Papers on Economic Activity, Fall*, pp. 261–97.

Gorton, Gary, and Andrew Metrick. 2012. "Securitized Banking and the Run on Repo." *Journal of Financial Economics* 104(3): 425–51.

Gorton, Gary B., Andrew Metrick, and Lei Xie. 2014. "The Flight from Maturity." NBER Working Paper 20027.

Government Accountability Office. 2009. *Financial Markets Regulation: Financial Crisis Highlights Need to Improve Oversight of Leverage at Financial Institutions and Across System*. Report to Congressional Committees. GAO-09-739, July. <https://www.gao.gov/new.items/d09739.pdf>.

Government Accountability Office. 2017. "Large Bank Supervision: Improved Implementation of Federal Reserve Policies Could Help Mitigate Threats to Independence." GAO-18-118, November. <https://www.gao.gov/assets/690/688140.pdf>.

Greenberger, Michael. 2010. "The Role of Derivatives in the Financial Crisis." Testimony, Financial Crisis Inquiry Commission, Washington D.C., June. https://fcic-static.law.stanford.edu/cdn_media/fcic-testimony/2010-0630-Greenberger.pdf.

Greenspan, Alan. 1997. "The Evolution of Banking in a Market Economy." Remarks at the Annual Conference of the Association of Private Enterprise Education, Arlington, Virginia, April 12. <https://www.federalreserve.gov/boarddocs/speeches/1997/19970412.htm>.

House of Representatives, Committee on Oversight and Government Reform. 2008. "The Financial Crisis and the Role of Federal Regulations." Hearing before the Committee on Oversight and Government Reform, House of Representatives, 110th Congress, 2nd Session,

October 23. Government Printing Office. <https://www.gpo.gov/fdsys/pkg/CHRG-110hhrg55764/html/CHRG-110hhrg55764.htm>.

IGM Forum. 2017. "Factors Contributing to the 2008 Global Financial Crisis." October 17, 2017. <http://www.igmchicago.org/surveys-special/factors-contributing-to-the-2008-global-financial-crisis>.

Inspector General of the Securities and Exchange Commission. 2008. *SEC's Oversight of Bear Stearns and Related Entities: The Consolidated Supervised Entity Program*. U.S. Securities and Exchange Commission, Office of the Inspector General, Report Number 446-A, September 25. <https://www.sec.gov/files/446-a.pdf>.

Inspector General of the Securities and Exchange Commission. 2009. "Investigation of Failure of the SEC to Uncover Bernard Madoff's Ponzi Scheme." US Securities and Exchange Commission, Office of the Inspector General, Report Number OIG-509, August. <https://www.sec.gov/news/studies/2009/oig-509.pdf>.

International Monetary Fund. 2007. "United States: Selected Issues." IMF Country Report 7/265, August. <https://www.imf.org/external/pubs/ft/scr/2007/cr07265.pdf>.

International Monetary Fund. 2010. "United States: Publication of Financial Sector Assessment Program Documentation—Technical Note on Consolidated Regulation and Supervision." IMF Country Report No. 10/251, Washington, July. <https://www.imf.org/external/pubs/ft/scr/2010/cr10251.pdf>.

Jackson, Thomas H. 2015. "Building on Bankruptcy: A Revised Chapter 14 Proposal for the Recapitalization, Reorganization, or Liquidation of Large Financial Institutions." In *Making Failure Feasible*, edited by Kenneth E. Scott, Thomas H. Jackson, and John B. Taylor. Hoover Institution Press.

Kacperczyk, Marcin, and Philipp Schnabl. 2010. "When Safe Proved Risky: Commercial Paper during the Financial Crisis of 2007–2009." *Journal of Economic Perspectives* 24(1): 29–50.

Kacperczyk, Marcin, and Philipp Schnabl. 2013. "How Safe Are Money Market Funds?" *Quarterly Journal of Economics* 128(3): 1073–1122.

Katz, Jonathan G. 2004. "Re: Proposed Rule: Alternative Net Capital Requirements for Broker-Dealers That Are Part of Consolidated Supervised Entities (Release 34-48690; SEC File No. S7-21-03)." March, 8. Securities and Exchange Commission. <https://www.sec.gov/rules/proposed/s72103/lehmanbrothers03082004.htm>.

Kohn, Don. 2014. "Institutions for Macroprudential Regulation: The UK and the U.S." *On the Record*, Brookings Institution, April 17.

Kotz, David. 2010. "Testimony of H. David Kotz, Inspector General of the Securities and Exchange

Commission, Before the Financial Crisis Inquiry Commission,” Securities and Exchange Commission, May 5. <https://www.sec.gov/about/offices/oig/reports/reppubs/2010/writtentestimonyofdaavidkotzbeforefinancialcrisisinquirycommission.pdf>.

Krishnamurthy, Arvind, Stefan Nagel, and Dmitry Orlov. 2014. “Sizing Up Repo.” *Journal of Finance* 69(6): 2381–2417.

Levitt, Arthur. 1998. “Testimony Before the Senate Committee on Agriculture, Nutrition, and Forestry, Concerning the Regulation of the Over-the-Counter Derivatives Market and Hybrid Instruments.” Securities and Exchange Commission, July 30. <https://www.sec.gov/news/testimony/testarchive/1998/tsty0998.htm#body9>.

Library of Congress. 2010. “Dodd-Frank Wall Street Reform and Consumer Protection Act,” Public Law 111-203, U.S. Statutes at Large, pp. 111–203, July 21. <https://www.gpo.gov/fdsys/pkg/PLAW-111publ203/pdf/PLAW-111publ203.pdf>.

Lo, Andrew W. 2012. “Reading About The Financial Crisis: A Twenty-One-Book Review.” *Journal of Economic Literature* 50(1): 151–78.

Macchiaroli, Michael. 2009. “Letter to Ms. Orice M. Williams Brown, Director, Financial Markets and Community Investment United States Government Accountability Office,” Securities and Exchange Commission, Washington, DC, in Financial Crisis Inquiry Report, Appendix. Reprinted in Government Accountability Office Report Number GAO-09-739 entitled “Financial Crisis Highlights Need to Improve Oversight of Leverage at Financial Institutions and Across System,” Washington, July. <https://www.gao.gov/assets/300/292767.html>.

McCaffrey, David P. 2016. “The Transformation of the Over-the-Counter Derivatives Market, 1984–2016.” November 29. Available as SSRN: <https://ssrn.com/abstract=2642216>.

McDonald, Robert, and Anna Paulson. 2015. “AIG in Hindsight.” *Journal of Economic Perspectives* 29(2): 81–106.

McLean, Bethany. 2012. “The Meltdown Explanation That Melts Away.” *Reuters*, March 19.

Meyer, Lawrence H. 2000. “Remarks Before a National Bureau of Economic Research Conference, Prudential Supervision: What Works and What Doesn’t.” Speech, Cheeca Lodge, Islamorada, Florida, January 14. Federal Reserve Board of Governors. <https://www.federalreserve.gov/boarddocs/speeches/2000/20000114.htm>.

Mian, Atif, and Amir Sufi. 2014. *House of Debt: How They (and You) Caused The Great Recession, and How We Can Prevent It From Happening Again*. University of Chicago Press.

Office of the Comptroller of the Currency. 2013. “An International Review of OCC’s Supervision of

Large and Midsize Institutions: Recommendations to Improve Supervisory Effectiveness.” December 4. Office of the Comptroller of the Currency, Washington. <https://www.occ.gov/news-issuances/news-releases/2013/nr-occ-2013-184a.pdf>.

Ohlrogge, Michael, and Kay Giesecke. 2016. “Can the SEC’s Net Capital Rule Really Explain The Growth of Mortgage Securitization?” May 2. http://www.web.stanford.edu/~ohlrogge/downloads/Rethinking_Securitization.pdf.

Polakoff, Scott M. 2009. “American International Group: Examining What Went Wrong, Government Intervention, and Implications for Future Regulation,” Statement of Scott M. Polakoff, Acting Director, Office of Thrift Supervision, Before the Banking, Housing, and Urban Affairs Committee of the United States Senate, March 5. Office of Thrift Supervision, Department of the Treasury. <https://occ.gov/static/ots/testimony/ots-testimony-ts171-03-05-2009.pdf>.

President’s Working Group on Financial Markets. 1999. “Over-The-Counter Derivatives Markets and the Commodity Exchange Act.” November. <https://www.treasury.gov/resource-center/fin-mkts/Documents/otcact.pdf>.

Rosengren, Eric S. 2013. “Bank Capital: Lessons from the U.S. Financial Crisis.” Federal Reserve Bank of Boston, Keynote Address, Bank for International Settlements Forum on Key Regulatory and Supervisory Issues in a Basel III World, Seoul, Korea, February 25. <https://www.bostonfed.org/news-and-events/speeches/bank-capital-lessons-from-the-us-financial-crisis.aspx>.

Rubin, Robert E., Alan Greenspan, and Arthur Levitt. 1998. “Letter dated June 5, 1998 to The Honorable Newt Gingrich, Speaker, U.S. House of Representatives, from Robert E. Rubin, Secretary, Department of the Treasury, Alan Greenspan, Chairman, Board of Governors of the Federal Reserve System, and Arthur Levitt, Chairman, Securities and Exchange Commission.” June 5. Press Center, U.S. Department of the Treasury. <https://www.treasury.gov/press-center/press-releases/Pages/cftcltr.aspx>.

Sarin, Natasha, and Larry H. Summers. 2016. “Understanding Bank Risk through Market Measures.” *Brookings Papers on Economic Activity*, Fall, pp. 57–120.

Schapiro, Mary L. 2010. “Testimony Concerning the Lehman Brothers Examiner’s Report Before the House Financial Services Committee.” April 20. US Securities and Exchange Commission. <https://www.sec.gov/news/testimony/2010/ts042010mls.htm>.

Schmidt, Lawrence, Allan Timmermann, and Russ Wermers. 2016. “Runs on Money Market Mutual Funds.” *American Economic Review* 106(9): 2625–57.

Schroth, Enrique, Gustavo Suarez, and Lucian A. Taylor. 2014. "Dynamic Debt Runs and Financial Fragility: Evidence From the 2007 ABCP Crisis." *Journal of Financial Economics* 112(2): 164–89.

Securities and Exchange Commission. 2004. "Alternative Net Capital Requirements for Broker-Dealers That Are Part of Consolidated Supervised Entities; Supervised Investment Bank Holding Companies; Final Rules." *Federal Register* Part II Securities and Exchange Commission, 17 CFR Parts 200 and 240, June 21. <https://www.sec.gov/rules/final/34-49830.pdf>.

Sirri, Erik R. 2009. "Remarks at the National Economists Club: Securities Markets and Regulatory Reform," Securities and Exchange Commission, April 9. <https://www.sec.gov/news/speech/2009/spch040909ers.htm>.

Sirri, Eric. 2010. "SEC Regulation of Investment Banks," Testimony before the Financial Crisis Inquiry Commission, May 5, Securities and Exchange Commission. Washington, DC. https://fraser.stlouisfed.org/files/docs/historical/fct/fct/fcic_testimony_sirri_20100505.pdf.

Spillenkothen, Rich. 2010. "Notes on The Performance of Prudential Supervision in the Years Preceding the Financial Crisis by a Former Director of Banking Supervision and Regulation at the Federal Reserve Board (1991 to 2006)." Notes for Financial Crisis Inquiry Commission, May. https://fcic-static.law.stanford.edu/cdn_media/fcic-docs/2010-05-31%20FRB%20Richard%20Spillenkothen%20Paper-%20Observations%20on%20the%20Performance%20of%20Prudential%20Supervision.pdf.

Stanton, Thomas H. 2009. "The Failure of

Fannie Mae and Freddie Mac and the Future of Government Support for the Housing Finance System." *Journal of Law and Policy* 18(1): 217–61.

Tucker, Paul. 2014. "Capital Regulation in the New World: The Political Economy of Regime Change." August. Available at: <http://paultucker.me/capital-regulation-in-the-new-world-the-political-economy-of-regime-change/>.

US Department of the Treasury. 2008. "Blueprint for a Modernized Financial Regulatory Structure." Washington, March 31. <https://www.treasury.gov/press-center/press-releases/Documents/Blueprint.pdf>.

US Department of the Treasury. 2018. "Report to The President of The United States: Orderly Liquidation Authority and Bankruptcy Reform." (Pursuant to the Presidential Memorandum Issued April 21.) February 21. https://home.treasury.gov/sites/default/files/2018-02/OLA_REPORT.pdf.

Valukas, Anton R. 2010. *Lehman Brothers Holdings Inc. Chapter 11 Proceedings Examiner Report*. Jenner & Block. March. <https://jenner.com/lehman>.

Vitan, Esmerelda. 2013. "Lehman's Demise and Repo 105: No Accounting for Deception." Knowledge@Wharton, May. <http://knowledge.wharton.upenn.edu/article/lehmans-demise-and-repo-105-no-accounting-for-deception>.

Yellen, Janet L. 2015. "Improving the Oversight of Large Financial Institutions." Speech, New York City, New York, Federal Reserve Board of Governors, March 3. <https://www.federalreserve.gov/newsevents/speech/yellen20150303a.htm>.

Would Macroprudential Regulation Have Prevented the Last Crisis?

David Aikman, Jonathan Bridges, Anil Kashyap,
and Caspar Siegert

A key response of official sectors around the world to the financial and economic crises of ten years ago has been the formation of financial stability committees. Such committees now exist in over 40 countries worldwide (Edge and Liang 2017). The remits of these committees are “macroprudential.” Macroprudential policy focuses on potential system-wide risks and amplification mechanisms, complementing the detailed firm-specific risk assessments of microprudential regulators. In addition, it has the explicit objective to ensure that the financial system does not amplify a downturn in the real economy—for example, by being forced to cut back on the supply of credit in a stress (Borio 2003; in this journal, Hanson, Kashyap, and Stein 2011).

This paper asks whether macroprudential authorities, as they have been designed over the past decade, could prevent—or materially dampen—a rerun of the last crisis. To be clear at the outset, macroprudential regulation does not seek to

■ *David Aikman is Technical Head of the Macroprudential Strategy and Support Division, Bank of England, London, United Kingdom. Jonathan Bridges is a Senior Economist, Macroprudential Strategy and Support Division, Bank of England, London, United Kingdom. Anil Kashyap is the Edward Eagle Brown Professor of Economics and Finance at the University of Chicago Booth School of Business, Chicago, Illinois. Since 2016, he has been a Member of the Financial Policy Committee, Bank of England. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts, and a Research Fellow, Centre for Economic Policy Research, London, England. Caspar Siegert is a Senior Economist, Macroprudential Strategy and Support Division, Bank of England, London, United Kingdom. Kashyap is the corresponding author at anil.kashyap@chicagobooth.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at

<https://doi.org/10.1257/jep.33.1.107>

doi=10.1257/jep.33.1.107

eliminate recessions. Instead, it is aimed at ensuring that the financial system does not create shocks that trigger recessions or amplify other shocks to make recessions materially worse. With this in mind, the first part of our paper provides an account of the amplifying factors that made the last crisis so severe. Our diagnosis centers on two overlapping but distinct vulnerabilities: the increase in leverage and short-term funding at financial intermediaries, and the build-up in indebtedness in the household sector. These factors, we argue, can account for around two-thirds to three-quarters of the fall in US GDP that followed the financial crisis. We describe and calibrate the policy interventions required to address these vulnerabilities.

We then contrast how well-equipped two prominent macroprudential regulators are to make these interventions. We argue that the US Financial Stability Oversight Council would likely make little difference were we to experience a rerun of the factors that caused the last crisis. It has no macroprudential levers under its direct control, and not all of its members have mandates to protect financial stability. A macroprudential regulator modeled on the UK's Financial Policy Committee stands a better chance because it has many of the necessary powers. But spotting build-ups in vulnerabilities in real-time is challenging. And given the role played by loosely regulated nonbank financial institutions prior to the last crisis—and the continuing evolution of the financial system—a successful macroprudential intervention would likely require political backing to be nimble in widening the perimeter of regulation to capture such institutions. More generally, such a regulator would have to be fairly aggressive in using its powers. Given the novelty of these powers, there is no clear evidence on whether such forceful interventions would be realistic were risks to escalate again. Our conclusion distills some key challenges and priorities for the development of a successful macroprudential framework.

The test we pose is really not very tough. Today's macroprudential frameworks were created in response to the scenario we are revisiting, whereas the challenges facing macroprudential regulators in the future will likely be new. But while our essay explores how today's macroprudential regimes might respond if vulnerabilities similar to those that caused the last crisis were to reoccur, we also invite readers to use this thought-experiment to consider how macroprudential committees might respond if other "resilience gaps" opened up in the future.

Fault-Lines That Led to the 2008 Financial Crisis

In Aikman, Bridges, Kashyap, and Siegert (2018), we describe the competing factors that contributed to the 2008 financial crisis and discuss the dimensions of their relative contributions. Here we summarize the evidence regarding the two dominant contributors: 1) the fragilities in the financial system associated with excessive leverage and the use of potentially flighty short-term funding; and 2) the unprecedented (by US standards) lending boom to the household sector that began in the mid-2000s. Bernanke (2018) also identifies these two channels to be of primary importance, with particular emphasis on the former.

Table 1

Size and Structure of the Leveraged Financial System

| | 2001:Q4 | | | | 2007:Q4 | | | |
|-------------------------------------|------------------|----------|------------------|-----------------------|------------------|----------|------------------|-----------------------|
| | Assets (\$bn) | Leverage | Liquid assets | Short-term funding | Assets (\$bn) | Leverage | Liquid assets | Short-term funding |
| Commercial banks | 6,552 | 11.0 | 6.6% | 26.5% | 11,182 | 9.8 | 4.6% | 33.2% |
| <i>of which: large institutions</i> | 2,291 | 12.2 | 6.7% | 32.9% | 5,422 | 11.8 | 4.6% | 37.5% |
| Savings institutions | 1,317 | 11.6 | 3.0% | 18.2% | 1,852 | 9.1 | 2.3% | 22.6% |
| Broker-dealers | 2,376 | 28 | 2.4% | 57.3% | 4,686 | 45 | 0.4% | 63.4% |
| Government-sponsored enterprises | 1,417 | 42.3 | 0.2% | | 1,677 | 23.7 | 0.7% | |
| Total | 11,662 | | | | 19,397 | | | |

Source: Financial Accounts of the United States; Call Reports; FDIC; Adrian, Fleming, Shachar, and Vogt (2017); and Annual Reports of Fannie Mae and Freddie Mac.

Note: "Leverage" is defined as total assets divided by (book) equity. "Liquid assets" refers to the ratio of cash and Treasury securities to total assets. For brokers, "short-term funding" refers to repo funding relative to total assets. For deposit-takers, it refers to (estimated) uninsured domestic deposits and foreign deposits relative to total assets. While deposits are typically short-term liabilities, many types of deposits, including insured deposits in particular, are "behaviorally stable" and were not withdrawn during the crisis (Martin, Puri, and Ufier 2018). "Large commercial banks" are defined as banks with at least \$150 billion in total assets. For 2007, this is adjusted using the Consumer Price Index (\$180 billion). Government-sponsored enterprises include Fannie Mae and Freddie Mac.

Our thesis is that these two factors amplified the initial losses that occurred when house prices fell. The fragilities in the financial system meant that lenders had to cut back lending as they struggled to absorb losses, which led to a credit crunch that reduced investment and employment. As households also struggled to deal with excessive debt, they cut spending, amplifying the downturn further.

Fragilities in the Financial System

Vulnerabilities in the financial system built significantly in the years leading up to the global financial crisis (Brunnermeier 2009; Acharya, Philippon, Richardson, and Roubini 2009; Duffie 2018). As a result, even relatively small losses on financial institutions' mortgage exposures were sufficient to trigger stability concerns for the entire financial system that ultimately spilled over into the real economy.

Table 1 documents the solvency, liquidity, and funding positions of different classes of US financial institutions at two points prior to the global financial crisis: end-2001, a period when the US economy was recovering from the strains caused by the collapse of the dot-com bubble; and end-2007, the beginning of the financial crisis. The total assets of the institutions we capture here increased from \$12 trillion to almost \$20 trillion between these dates. Clearly, the use of debt finance, or "leverage," varied significantly across institutions. For the largest commercial banks, leverage changed little in the years leading up to the crisis. Commercial banks did, however, reduce the amount of assets that could easily be sold without price

concessions (liquid assets) and expanded their reliance on short-term funding, which can rapidly disappear during times of stress.

The most extreme vulnerabilities developed for the parts of the financial system that did not take traditional deposits. Consider, for instance, the changes for broker-dealers, a category that includes specialised investment banks and the investment banking subsidiaries of larger banking groups. The assets of these entities increased from 28 to 45 times their equity between 2001 and 2007, meaning that a roughly 2 percent decline in the value of broker-dealers' assets would have been sufficient to wipe out all of their equity. In addition, these firms were traditionally highly reliant on short-term wholesale funding (Rosengren 2014), and became even more so during this period.

Much of this short-term funding took the form of repurchase agreements, or "repos." Repos are a form of borrowing in which the broker-dealer sells securities that it holds, receives the value of those securities in cash, and a few days later repurchases the securities at a predetermined price that includes an additional interest payment. The repo liabilities of broker-dealers increased from \$1.4 trillion in 2001 to \$3.0 trillion in 2007.¹ Moreover, an increasing fraction of repos were backed by low-quality securities.

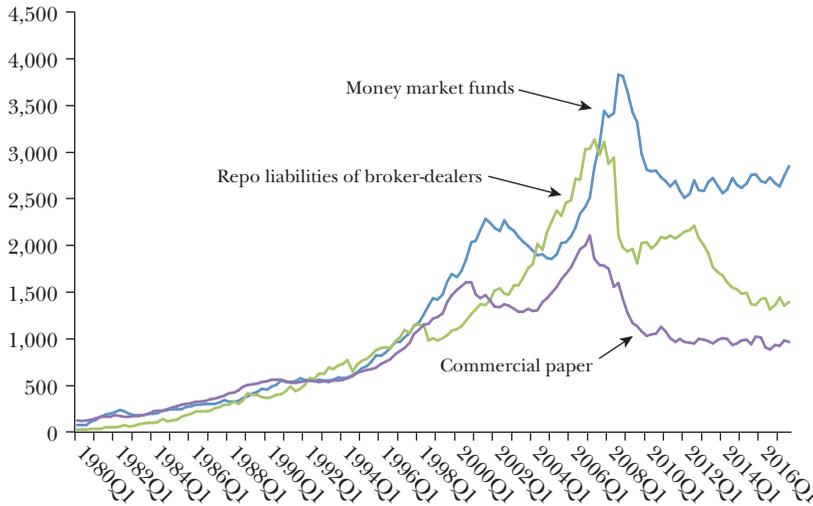
Figure 1 shows the rise in repo funding, along with commercial paper, another form of funding that experienced rapid growth over this period. Traditional commercial paper is short-term debt issued by companies to fund operations. However, by the end of 2006, 60 percent of outstanding commercial paper consisted of so-called "asset-backed commercial paper" that had been issued to fund the purchase of specific securities such as credit card receivables, auto loans, or mortgage-backed securities.

The growth in repos and commercial paper coincided with an increase in the size of money market mutual funds, which purchased much of the repos and commercial paper issued. Regulators allowed money market mutual funds to invest in assets with a weighted average maturity of up to 90 days, but these funds offered investors the ability to withdraw their money at a day's notice. Moreover, money market mutual funds did not have any capital that would shield these short-term investors from losses. In a crisis, investors in money market mutual funds who withdrew their funds first were certain to be fully paid, while later claims might not be fully paid, providing incentives to "run" on the fund.

In summary, nonbanks became an increasingly important source of credit for the real economy in the years preceding the crisis: between 2001 and 2007, nonbank financials accounted for over 70 percent of the total growth in home mortgage credit (according to the Financial Accounts of the United States). This growth was accompanied by an increased reliance on debt financing of the nonbank system. Short-term borrowing became more important, with the belief that it could be

¹ Total repo liabilities for all types of institutions recorded in the Financial Accounts of the US data for end-2001 and 2007 were \$2.2 trillion and \$4.8 trillion, respectively. None of these numbers were readily available in the run-up to the crisis, as broker-dealers repo liabilities were only reported on a netted basis (Eichner, Kohn, and Palumbo 2013; Holmquist and Gallin 2014).

Figure 1
Reliance on Short-Term Funding
 (\$ billions)



Source: March 2018 release of the Financial Accounts of the United States, based on Adrian, de Fontnouvelle, Yang, and Zlate (2017). The size of money-market funds is measured as outstanding money market fund shares (liabilities) in table L.121. Commercial paper refers to commercial paper (liabilities) issued by any sector (table L.2019), which includes asset-backed commercial paper. Repo liabilities of broker-dealers are based on security repurchase agreements (liabilities) in table L.130.

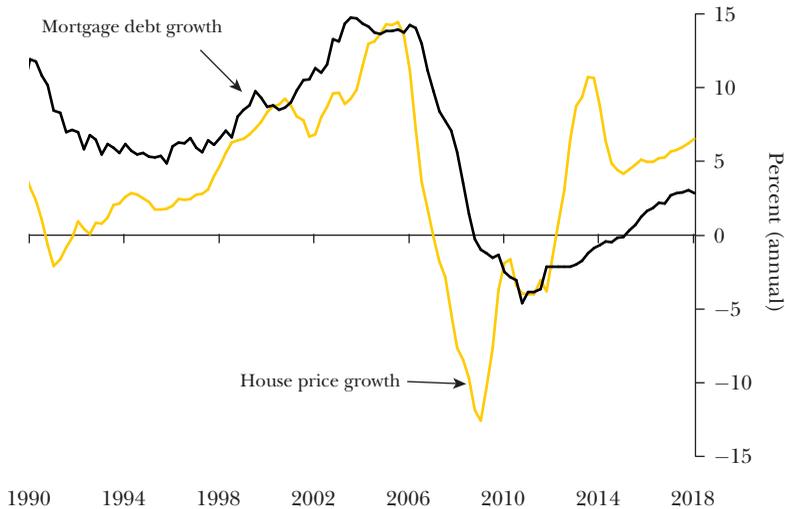
rolled over continually. These observations suggest that macroprudential regulators’ arsenals must include tools to affect the overall propensity to rely on debt financing and also affect the maturity of the funding, and macroprudential regulators must have scope to apply these tools both to banks and to nonbanks.

The Build-up in Household Debt

Alongside the pronounced build-up in leverage and short-term funding in the financial system, there was a rapid build-up in debt in the real economy, concentrated in household mortgages. Mortgage debt doubled in the six years before the crisis, and by 2007 reached 72 percent of GDP. Two aspects of this debt build-up are noteworthy and will inform our later macroprudential analysis.

First, the increase in mortgage debt was accompanied by a house price boom, shown in Figure 2. House prices rose by two-thirds in the five years to their peak in early 2006 (according to the S&P Case-Shiller US National Home Price Index), and ongoing rapid house price appreciation was embedded in expectations (Gennaioli and Shleifer 2018). The aggregate loan-to-value ratio on the stock of US housing remained broadly flat during this period, meaning that for each 1 percent increase in house values, homeowners also increased their mortgage debt by around 1 percent. In part, this reflected new homeowners taking out larger mortgages in order to purchase more expensive homes. But in addition, existing homeowners

Figure 2
Mortgage Debt and House Prices
(annual percentage change)



Source: Financial Accounts of the United States and S&P Case-Shiller US National Home Price Index.

also extracted housing equity by taking out additional debt. Mian and Sufi (2011) estimate that existing homeowners borrowed \$0.25 on average for every \$1 increase in home-equity value during the housing boom, enough to account for over half of the increase in debt for homeowners between 2002 and 2006.

Second, there were clear signs in the years before the financial crisis that lending standards were being loosened and borrower quality was deteriorating. The Federal Reserve Board's Senior Loan Officer Opinion Survey on Bank Lending Practices reported easing standards between 2004Q1 and 2006Q3. The expansion of credit to the most risky borrowers was particularly pronounced. For example, according to the Federal Reserve's Survey of Consumer Finances, the share of the stock of mortgages with debt of over four times their income more than doubled between 2001 and 2007 from 6 percent to 13 percent. The number of new subprime mortgages nearly doubled between 2003 and 2005, 80 percent of which were made with short-term "teaser" interest rates (in this journal, Mayer, Pence, and Sherlund 2009). "Near-prime" mortgages also increased rapidly. The private-label securitization market, in which these mortgages were bundled into tranching financial securities and resold, was an important driver of these frothy credit supply conditions (Keys, Mukherjee, Seru, and Vig 2010).

In summary, the years running up to the Great Recession saw an unprecedented surge in US household debt. That boom was accompanied and reinforced by soaring property prices. Aggressive credit supply expansion, compounded by financial innovation, provided the undercurrent for an unsustainable cycle. Household balance sheets became increasingly vulnerable to a shock as more credit was extended to highly indebted households.

Quantifying the Effects of the Credit Crunch and Deleveraging

Extrapolating the 20-year average growth rate from 2007Q4 through 2010Q4 suggests that the level of GDP per capita was 8.5 percent below trend by 2010Q4. How much of this might plausibly be attributed to the credit crunch and the deleveraging by overly indebted households? Estimating precise contributions is challenging. Nonetheless, triangulating across a range of studies, it seems very likely that these factors account for a large part of that gap. In Aikman, Bridges, Kashyap, and Siegert (2018), we provide the details of how we arrive at this conclusion; here, we simply summarize the main points.

A variety of studies, using a variety of methods, find that the fragility of lenders resulted in a credit crunch that had the potential to materially affect real economic activity (for example, Chodorow-Reich 2014; Greenlaw, Hatzius, Kashyap, and Shin 2008; Bassett, Chosak, Driscoll, and Zakrajšek 2014; Guerrieri et al. 2015). Translating the estimates from any one of these studies into an impact on the GDP shortfall requires a number of assumptions. For example, Chodorow-Reich (2014) identifies the impact of the credit crunch on employment in enterprises with less than 1000 employees. For our purposes, this finding needs to be extrapolated to the entire economy and then translated into an impact on GDP. Averaging across our five preferred studies, in Aikman, Bridges, Kashyap, and Siegert (2018), we find that about 35 percent of the 2010 GDP gap can be attributed to the abrupt tightening of credit conditions. That is, around 3 percentage points of the overall GDP shortfall can be explained by the fragilities in the financial system, which meant that the economy was prone to suffering a credit crunch.

There is also convincing evidence that a strong relationship exists between household debt growth in the years preceding economic downturns and the severity of the subsequent downturn (for example, Jordà, Schularick, and Taylor 2013, 2016; Bridges, Jackson, and McGregor 2017; Mian and Sufi 2012; Mian, Sufi, and Verner 2017; Gertler and Gilchrist 2018). Converting the various estimates into an impact on the fall in GDP during the Great Recession is subject to the same basic challenges as with scaling the impact of the credit crunch. In Aikman, Bridges, Kashyap, and Siegert (2018), we average across a number of approaches and find that the household debt boom can account for about one-half of the GDP gap, or just over 4 percentage points of the overall fall in GDP.

Our reading of the existing evidence is therefore that, taken together, these two effects account for around three-quarters of the contraction in output that occurred during the Great Recession. That is, absent the credit crunch and the deleveraging by households, the cumulative fall in GDP growth during the recession would have been three-quarters smaller.

It is obviously an oversimplification to treat the effects of the credit crunch and the household deleveraging channels as if they were clear and distinct events. Rather, there was two-way feedback between these phenomena: tight credit conditions intensified households' need to deleverage, and the reduction in spending by highly indebted households led to an economic contraction that made it harder for all borrowers to service their debts, generating larger losses for banks and other

financial intermediaries. However, disentangling the precise impact of these channels is not material to our argument. Rather, our claim is that each channel had a very substantial bearing on the costs of the crisis. Should we see a rerun of the factors that caused the last crisis, therefore, macroprudential policy would have to address vulnerabilities associated with excessive debt-financing and short-term funding in the financial system, and excessive debt levels in the household sector.

What Could a Macroprudential Regulator Have Done to Address the Build-Up in These Vulnerabilities?

We start this section by asking whether it was possible to spot the vulnerabilities in both the financial system and in household balance sheets documented above in real time. We then consider the policies a macroprudential regulator could introduce in response. In the next section, we discuss the institutional frictions that actual macroprudential regulators would face in implementing such policies in practice.

Identifying the Build-up in Risk in Real Time

Identifying macroprudential policy interventions requires spotting emerging risks and accumulating vulnerabilities prior to a crisis. Identifying the debt build-up in the household sector was relatively straightforward. The Bank for International Settlements was sounding alarms about the risks from credit build-ups in 2004 (for example, Borio and White 2004). The IMF's *Global Financial Stability Report* in April 2005 had a chapter on the state of household balance sheets in advanced economies.

Table 2 summarizes how often some key words associated with building fragilities were mentioned in Federal Open Market Committee (FOMC) transcripts through the 2000s. Again, the build-up in household debt was clearly evident from the early 2000s. The house price bubble was also observed well ahead of the crisis. By its meeting in June 2005, the FOMC was discussing evidence that houses may be up to 20 percent overvalued, leading to the spike in transcript references to “house price,” “bubble,” and words associated with mortgage lending. This assessment turned out to be pretty accurate: by the end of the recession in June 2009, house prices were 13 percent below their June 2005 level, and the peak-to-trough fall during the crisis was 20 percent (based on the Case-Shiller US National Home Price Index). Between the end of the recession in June 2009 and February 2012, house prices fell further, bringing the overall peak-to-trough fall to 26 percent. However, the extent to which the build-up in debt was being concentrated at riskier, heavily indebted borrowers was not being adequately picked up (Eichner, Kohn, and Palumbo 2013). It is striking that the word “subprime” was mentioned 314 times in the FOMC's 2007 transcripts, but only 27 times in all the transcripts from 2000 to 2006. Commercial paper and “securitization” were also rarely mentioned before 2007.

More broadly, policymakers did not understand the effects that a sharp fall in house prices would have on the financial system. This lack of resilience might have been identified via stress tests—since the crisis, such tests have become a key

Table 2

Financial Stability Terms Appearing in Discussions of the Federal Open Market Committee

| | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|--|-------|------|-------|-------|------|-------|-------|
| <i>General</i> | | | | | | | |
| “Financial stability” | 0 | 1 | 2 | 9 | 5 | 9 | 13 |
| <i>Financial System:</i> | | | | | | | |
| “Bank” | 502 | 429 | 449 | 302 | 284 | 309 | 1,024 |
| “Capital”/“Leverage” | 454 | 308 | 340 | 208 | 183 | 177 | 402 |
| “Shadow”/“Broker”/“Money market” | 17 | 21 | 40 | 10 | 17 | 28 | 59 |
| “Fund”/“Liquid”/“Repo” | 1,226 | 962 | 1,150 | 1,058 | 932 | 1,110 | 1,779 |
| “Commercial paper”/“Securitization” | 23 | 22 | 15 | 3 | 14 | 2 | 133 |
| <i>Housing Market</i> | | | | | | | |
| “House price” | 2 | 23 | 4 | 41 | 160 | 85 | 83 |
| “Bubble” | 6 | 15 | 14 | 19 | 114 | 4 | 8 |
| “Loan”/“Lend”/“Debt”/“Credit”/“Borrow” | 413 | 442 | 452 | 269 | 409 | 251 | 1,563 |
| “Mortgage” | 84 | 100 | 96 | 67 | 176 | 118 | 481 |
| “Subprime” | 0 | 3 | 1 | 0 | 8 | 15 | 314 |
| “LTV”/“Heloc”/“Teaser”/“Alt-A” | 2 | 1 | 1 | 0 | 40 | 0 | 45 |

Note: For each year, transcripts of the eight FOMC meetings and any Conference Calls were searched. All transcripts available here: https://www.federalreserve.gov/monetarypolicy/fomc_historical_year.htm. A simple count of all words containing the stem words listed in the table above was conducted.

component of macroprudential regulators’ toolkits. Hirtle, Kovner, Vickery, and Bhanot (2016) demonstrate that a top-down stress test based on a macroeconomic scenario like the one that played out in 2007–08 would have predicted a significant capital shortfall in the US banking system as early as 2004. But even with a severe stress test, it would have been difficult to understand the fragility of funding flows across the system prior to the crisis, which led to fragilities in the nonbank sector and amplified the macroeconomic downturn. To reveal the full extent of the vulnerabilities that existed, stress tests would have had to cover the entire financial system: broker-dealers; commercial paper, repo, and derivative markets; specialized investment vehicles (SIVs); and other conduits. Building a complete map of funding interconnections between these markets and entities is challenging even today.

Thus, while some warning signs were clearly present in the lead-up to the crisis, we are cautious about the ability of macroprudential regulators to understand the nature of systemic financial risks as they emerge in real time. One implication is that policymakers should seek to develop systematic frameworks within which to monitor emerging risks and their potential implications for macroeconomic tail events.² Another implication is that we should be humble (see also Tarullo 2014). The fact that even in hindsight we believe it would have been hard to diagnose fully the risks in the run-up to the crisis suggests that macroprudential policy frameworks

² Analysis of “GDP-at-risk” and its link to financial indicators is one promising avenue here—see for example Adrian, Boyarchenko, and Giannone (2016), IMF (2018b), and Aikman et al. (2018).

should be calibrated with some built-in “slack” to account for the inherent difficulty of risk assessment, particularly in real time.

Tools and Actions to Reduce Leverage

How much additional capital would US banks have needed to be resilient given the extent of the credit bubble that was building in this period? For a sense of the necessary scale, consider the government capital support that occurred at the height of the crisis via the Troubled Asset Relief Program (TARP). Under this scheme, the US Treasury invested \$200 billion in the preferred stock of 15 large US banks to enhance market confidence in the banking system and to increase its capacity to lend.³ While establishing cause and effect is difficult, there is evidence that this intervention led to dramatic improvements in how market participants viewed the solvency of the US banking system. For example, the interest rate spreads for banks’ unsecured borrowing—often measured by looking at the difference between the three-month interbank borrowing rate and the risk-free Treasury bill yield—fell sharply almost immediately after the TARP was announced on October 14, 2008.

One means by which the authorities could increase system-wide levels of capital in the banking system in response to an emerging “resilience gap” is through a macroprudential tool called the countercyclical capital buffer (CCyB). The CCyB allows regulators to increase capital requirements according to the aggregate risk environment. What level of the countercyclical capital buffer would have delivered a level of resilience equivalent to the TARP injections? The countercyclical capital buffer is typically expressed as a percentage of a firm’s assets, weighted by the riskiness. It is then adjusted by a “domestic lending conversion factor,” which accounts for the fact that large banks operate across international boundaries. An estimate of the necessary countercyclical capital buffer rate is hence:

$$\begin{aligned} &\text{Required countercyclical capital buffer} \\ &= \frac{\$200 \text{ billion}}{\text{Risk-weighted assets}} * \text{Domestic lending conversion factor} \end{aligned}$$

As of 2005, the 15 TARP recipients on which we focus had total risk-weighted assets of approximately \$8.4 trillion—the denominator of the expression above.⁴ The average “domestic lending conversion factor” was around 75 percent (Avraham, Selvaggi, and Vickery 2012). That is, because large US banks have substantial global assets,

³ The 15 bank holding companies and broker-dealers that received the largest injections in the Troubled Asset Relief Program (TARP) were Citigroup, Bank of America, JP Morgan Chase, Wells Fargo, Goldman Sachs, Morgan Stanley, PNC Financial Services Group, U.S. Bancorp, SunTrust, Capital One Financial, Regions Financial Corporation, Fifth Third Bancorp, BB&T, Bank of New York Mellon, and Key Corp. The estimates we report in the text do not include capital provided by this program to other, smaller banks.

⁴ This number is estimated using published accounts and an average risk-weight of 67.5 percent (based on the New York Fed Quarterly Trends for Consolidated U.S. Banking Organizations). This total includes the assets of firms that did not themselves receive TARP assistance, but that in the course of 2008 were acquired by one of the 15 TARP-recipients on which we focus. These acquired firms include Countrywide Financial, Merrill Lynch, Wachovia, Bear Stearns, Washington Mutual, and National City Corp.

Table 3

Countercyclical Capital Buffer Rate (CCyB) That Would Have Been Necessary to Avoid the Troubled Asset Relief Program

| <i>Calculation</i> | | |
|--|--|-----------|
| Baseline: Replacing bail-outs | | |
| Total capital injections | | \$198bn |
| Total risk-weighted assets (RWAs) | | \$8,409bn |
| Bailout in percent of RWAs | \$198bn/\$8,409bn | 2.4% |
| Domestic assets in percent of total assets | | 76% |
| Required CCyB rate | 2.4%/76% | 3.1% |
| Variant 1: Replacing bail-outs and private sector capital raising | | |
| Additional private sector capital raising | | \$70bn |
| Required CCyB rate | $3.1\% \times (\$198\text{bn} + \$70\text{bn}) / \$198\text{bn}$ | 4.2% |
| Variant 2: Replacing bail-outs, and supporting additional lending | | |
| Additional RWAs if credit growth had continued along pre-crisis trend | | \$1,050bn |
| Assumed stressed target capital ratio | | 10% |
| Additional capital to support credit growth | \$1,050bn \times 10% | \$105bn |
| Required CCyB rate | $3.1\% + \$105\text{bn} / (\$8,409\text{bn} \times 76\%)$ | 4.7% |

Source: US Treasury, Published Accounts; New York Fed Quarterly Trends for Consolidated US Banking Organizations; Financial Accounts of the United States; Avraham et al. (2012).

Note: For variant 2, we assume that banks balance sheets had grown by 7 percent rather than 1 percent per year over two consecutive years. This is in line with the difference in the commercial bank credit growth rate between the 20 years before the crisis and the crisis (Q4 2007 to Q4 2009).

an increase in the US countercyclical capital buffer rate will not pass through one-for-one into their capital requirements. Using these parameters and the calculations in Table 3, we estimate that a countercyclical capital buffer of 3 percent would have provided an equivalent level of resilience as the \$200 billion TARP injection. Had a countercyclical capital buffer of 3 percent been built-up in the run-up to the crisis, it would have, in effect, brought the capital raising that ultimately proved necessary forward in time, substituting public provision of capital for private sector resources.

The approach above does not account for the capital that banks raised privately after the results of the stress tests through the Supervisory Capital Assessment Program were published in May 2009. Banks had six months to raise any required capital in private markets, with an explicit backstop option to obtain capital from the US Treasury if necessary. The banks in our sample raised approximately \$70 billion of capital in order to meet these requirements and did not turn to additional government funding. If this \$70 billion is added to the \$200 billion of TARP, our thought experiment suggests a countercyclical capital buffer of around 4.2 percent would have been required to bring forward the public and private capital raising that occurred during the heights of the crisis.

While TARP significantly reduced stress in the banking system, it was not fully sufficient to restart the provision of credit to the economy.⁵ Thus, a second

⁵ Chavaz and Rose (forthcoming) show that the effect of TARP on the provision of mortgage credit differed across regional markets and that TARP recipients reduced mortgage lending in the average county.

sensitivity check is to estimate what size of the countercyclical capital buffer would have allowed banks to continue lending in line with historical credit growth rates. As shown by the calculations in Table 3, a countercyclical capital buffer of 4.7 percent would have ensured that banks would have had sufficient capital to avoid applying for TARP and to continue growing their balance sheets in line with the long-run average growth rate.

Given its profitability in the years preceding the crisis, the banking system had ample capacity to meet increases in the countercyclical capital buffer rate of this magnitude through a combination of new issuance of equity and additional retentions. Hirtle (2016) finds that between 2005 and the collapse of Lehman Brothers in 2008, dividend payments of large bank holding companies amounted to \$162 billion, and total share buy-backs amounted to an additional \$131 billion. Indeed, dividend payments and share buy-backs amounted to \$49 billion and \$18 billion, respectively, between mid-2007 and the failure of Lehman in September 2008. By mid-2007, New Century Financial Corporation, a leading subprime mortgage lender, had already failed, and Bear Stearns and BNP Paribas had started halting redemptions on a number of their investment funds.

Finally, we note that among the 15 institutions that we consider are some broker-dealers. These institutions were not subject to standard prudential requirements.⁶ As a first step, a macroprudential authority would have needed to bring these firms inside the regulatory perimeter. As illustrated in Table 1, bringing all US broker-dealers to the same capital standards that commercial banks had in 2007 would already have added a substantial amount of capital to the system.

Tools and Actions to Reduce Funding Mismatches

During the financial crisis, the Federal Reserve set up a number of new liquidity facilities (Fleming 2012).⁷ These facilities, which were phased out within a few years of the end of the crisis, provided around \$1.5 trillion of short-term liquidity to the financial system, an amount equivalent to 9 percent of commercial banks' and broker-dealers' assets. We posit that requiring firms to replace \$1.5 trillion of their short-term funding with longer-term debt would have reduced liquidity outflows in the crisis in a way that would have avoided a need for extraordinary central bank liquidity facilities. This is likely to be an overestimate of the scale of appropriate policy intervention because some public provision of liquidity in a crisis is likely to be efficient (Holmström and Tirole 1998).

⁶ In 2004, the Securities and Exchange had created the voluntary "Consolidated Supervised Entities" program to regulate large investment bank holding companies. However, this regime was primarily intended to satisfy foreign regulators (Financial Crisis Inquiry Commission 2011). It was generally seen as being insufficiently robust and was terminated following the failure of Lehman Brothers in 2008.

⁷ The facilities included the Discount Window Funding, the Term Auction Facility, the Primary Dealer Credit Facility, the Term Securities Lending Facility, the Term Asset-Backed Securities Loan Facility, the Commercial Paper Funding Facility, and the Asset-Backed Commercial Paper Money Market Mutual Fund Liquidity Facility.

As a method of reducing the risks that entail when banks fund long-term, illiquid assets with short-term funding, the Basel Committee on Banking Supervision proposed a “net stable funding ratio” standard that took effect in 2018 (for an overview see, <https://www.bis.org/fsi/fsisummaries/nsfr.pdf>). Data provided in Lallour and Mio (2016) suggest that prior to the crisis, applying the net stable funding ratio to 12 of the largest US banking and investment banking groups at a consolidated level would have led to an increase in long-term funding of \$1.4 trillion by end-2007.⁸

What impact might such an intervention have on the real economy? While a \$1.5 trillion increase in the supply of long-term debt would probably affect equilibrium yields, we think such an intervention would increase firms’ average funding costs by less than 10 basis points. Conservative estimates suggest this in turn might increase lending spreads by less than 20 basis points, reducing the level of GDP by less than 0.2 percent.⁹

Tools and Actions to Reduce Build-Up in Household Debt

Higher capital and liquidity requirements might also reduce household debt growth and house prices by increasing the cost of credit for borrowers. However, the impact of implementing such measures in a boom may be small (for evidence, see Bahaj, Bridges, Malherbe, and O’Neill 2016). Thus, a macroprudential regulator determined to reduce a rapid build-up in household debt might wish to take additional actions.

Here, we consider the potential impact on the household debt boom of imposing loan-to-income limits and accompanying affordability criteria on new mortgages. Adelino, Schoar, and Severino (2016) document the widespread nature of the household credit boom: mortgage originations—and subsequent delinquencies—increased across the income distribution and across credit scores. By constraining unsustainable borrowing choices across the spectrum, macroprudential loan-to-income limits and affordability tests would therefore have helped to reduce the build-up of household debt vulnerabilities in the run-up to the crisis. We focus on these potential interventions rather than minimum down payment (“loan-to-value”) restrictions because the impact of the latter may have been limited by the twin nature of the household debt and house price booms. As discussed above, the aggregate loan-to-value ratio on the stock of US housing remained broadly flat

⁸ The estimates in Lallour and Mio (2016) are based on end-2006 balance sheet data. To make them comparable to the size of the Fed’s liquidity interventions, we scaled them up to reflect the average growth of the relevant groups’ balance sheets between 2006 and 2007.

⁹ The estimate of 10 basis points is based on the conservative assumption of a 100 basis point difference between the spreads on short-term funding and long-term (five-year) debt. To put this into context, the average difference between the cost of repo funding and five-year corporate bond spreads in 2006 was around 70 basis points. The estimated impact on lending spreads is based on the assumption that the increase in funding costs is fully passed on to borrowers by increasing spreads on loans (which represented about 50 percent of total assets), and that financial institutions’ cost of equity remains the same despite the more stable funding base. Given that around 10 percent of banks’ liabilities had to be replaced by long-term debt, this translates into an increase in lending spreads of 20 basis points. The impact of higher lending spreads on GDP is estimated based on multipliers in Firestone, Lorenc, and Ranish (2017).

Table 4

Impact of Different Loan-to-Income Limits on Gross Mortgage Lending for Owner-Occupier House Purchase (First Lien Loans Only)

| | Loans granted (number, millions) | Number of loans (millions) impacted by loan-to-income limit of: | | | | Loans granted (value, \$ billions) | Impact on value (\$ billions) of mortgages originated assuming all impacted loans reduced in size: | | | |
|--------------------------|----------------------------------|---|------------|------------|------------|------------------------------------|--|------------|-----------|-----------|
| | | 2x | 3x | 4x | 5x | | 2x | 3x | 4x | 5x |
| 2003 | 4.1 | 2.9 | 1.3 | 0.4 | 0.1 | 755.8 | 189.6 | 59.0 | 16.1 | 6.0 |
| 2004 | 4.6 | 3.2 | 1.5 | 0.5 | 0.1 | 906.6 | 245.8 | 81.9 | 21.7 | 6.1 |
| 2005 | 4.8 | 3.4 | 1.6 | 0.5 | 0.1 | 1,031.5 | 288.6 | 95.8 | 23.7 | 6.0 |
| 2006 | 4.2 | 2.9 | 1.4 | 0.4 | 0.1 | 939.5 | 245.9 | 75.0 | 18.7 | 6.0 |
| 2007 | 3.4 | 2.4 | 1.2 | 0.4 | 0.1 | 755.9 | 204.4 | 67.8 | 17.7 | 4.8 |
| Cumulative total: | | | | | | | | | | |
| 2003–2007 | 21.1 | 14.8 | 7.0 | 2.2 | 0.6 | 4,389 | 1,174 | 379 | 98 | 29 |

Source: Home Mortgage Disclosure Act (HMDA) data.

Note: The left panel identifies the *number* of mortgage originations for owner-occupier house purchase that would have been affected by loan-to-income limits set at the levels labelled. The right panels give the *value* reduction in gross lending that would have resulted if all those affected mortgages were reduced in size such that they just met the listed loan-to-income limit.

in the run-up to the crisis, whereas household debt rose sharply relative to income. Adelino, Schoar, and Severino (2017) also document the relatively stable distribution of combined loan-to-value ratios *at origination* between 2001 and 2007.¹⁰

As a simple illustration, consider how loan-to-income limits would have affected loan-level mortgage originations for owner-occupier house purchases in the run-up to the crisis.¹¹ For example, a loan-to-income limit of four times income applied from 2003 to 2007 would have meant that 2.2 million of the 21 million mortgages originated would have had to be reduced in size (see Table 4). Assuming all of these affected loans were still originated at the largest size possible given the limit, this intervention would have left the mortgage stock on the eve of the crisis around \$100 billion or about 1 percent lower than the \$10.6 trillion observed.

However, this naive experiment is likely to understate significantly the effect of loan-to-income limits on the mortgage stock. First, the data sample excludes second lien or “piggyback” mortgages, whereas well-designed loan-to-income limits would take into account the *combined* value of first and additional loans. To the extent that these additional loans pushed households’ overall indebtedness above the loan-to-income limit, some would have been curtailed. This could have had a material

¹⁰ Mayer, Pence, and Sherlund (2009) document an increase in the median combined loan-to-value ratio for *non-prime* purchase loans between 2003 and 2007. We argue below that income and affordability limits would have been effective in moderating the boom in non-prime mortgage lending.

¹¹ With thanks to Matthieu Chavaz for assistance, we use annual data resulting from the 1975 Home Mortgage Disclosure Act (HMDA), which covers the vast majority of mortgages originated. From this dataset, we analyse first lien mortgages for house purchase by owner-occupiers where both loan size and income is reported.

effect: the total flow of second lien mortgage originations from 2003 to 2007 was over \$200 billion. Second, this experiment focuses only on owner-occupier house purchase loans, with no impact assumed on investor loans or refinancing, both of which were important features of the household credit boom. Third, our calculation assumes that all affected borrowers still receive a loan and at the largest size possible given the limit—in reality, some borrowers would likely be shifted further below the limit and some might be excluded altogether. As an upper bound of the impact, if all originations constrained by a loan-to-income limit of four times income were excluded altogether—rather than just reduced in size—the impact on gross lending would rise from around \$100 billion to \$620 billion.

Borrowers at the riskier end of the spectrum that could either not certify their income or that had particularly stretched affordability characteristics would have perhaps been the most likely to have been excluded altogether. We can attempt to quantify these considerations. Between 2003 and mid-2007, about half of non-prime mortgage originations for house purchase had low or no documentation of income, assets, or both (Mayer, Pence, and Sherlund 2009). That amounted to around 1.7 million loans or 8 percent of the total captured in Table 4 over the same period.¹² These loans performed significantly worse than those with full documentation—by 2008 serious delinquencies on low or no documentation subprime loans had risen to 25 percent (Mayer, Pence, and Sherlund 2009). Many of these loans would likely have been curtailed by a loan-to-income limit, given the requirement that some actual documentation of income is provided in order to implement a loan-to-income policy. To give a sense of scale, had all of these loans been excluded by a loan-to-income policy between 2003 and mid-2007, lending could have been reduced by around \$360 billion (next to last line of Table 5).¹³

As a complement to loan-to-income limits, a macroprudential authority aiming to enhance the resilience of household sector balance sheets could also recommend the introduction of affordability tests, which require lenders to assess borrowers' capacity to service debts in different circumstances (as discussed in Bank of England 2017). Evidence from the Survey of Consumer Finances suggests that by 2007, 20 percent of the total stock of mortgagors had debt service burdens of over 40 percent of their income—a situation that might reasonably have been flagged by an affordability test. The impact of such tests on the rapid growth in non-prime mortgage borrowing could have been significant. For example, 76 percent of subprime mortgages for house purchase between 2003 and mid-2007 were short-term hybrid

¹²The Mayer, Pence, and Sherlund (2009) sample is based on data from First American Loan Performance. These data capture the vast majority of securitized, non-prime (that is subprime or near-prime “Alt-A”) first lien mortgage originations. The sample covers a total of 9.7 million originations, including investor loans and refinances. For comparability with Table 4, we focus on the subsample of about 3.4 million loans to owner-occupiers for house purchase. When comparing to Table 4, this subsample underestimates the total share of non-prime loans, since it does not capture mortgages retained by the lender rather than securitized.

¹³ This assumes low- or no-documentation loans had a proportionate share in the total value of originations in Table 4.

Table 5

Potential Impact of a 4× Loan-to-Income Limit and Accompanying Affordability Test on Household Debt Boom

| | |
|--|------------|
| Mortgage debt stock | |
| Total mortgage debt stock (2007) ^a | \$10,638bn |
| Gross flow of new mortgages (for owner-occupier house purchase) | |
| Total value of loans granted (2003 to 2007) ^b | \$4,389bn |
| Direct impact of 4× loan-to-income limit (2003 to 2007)^b | |
| Lower-bound estimate: all loans still originated at maximum size within limit: | - \$98bn |
| Upper-bound estimate: all loans with loan-to-income > 4× excluded altogether: | - \$622bn |
| Potential upper-bound impacts on non-prime lending (2003 to 2007)^c | |
| If income requirement excluded all low- or no-documentation subprime loans | - \$359bn |
| If affordability test excluded all non-prime originations on teaser rates | - \$366bn |

^a Financial accounts of the United States.

^b Home Mortgage Disclosure Act (HMDA) data; gross flow of first lien owner-occupier purchase loans. Impact estimates do not include potential reduction in second lien loans, investor loans, or loans for refinance.

^c Number of non-prime first lien owner-occupier purchase loans estimated based on Mayer et al. (2009) and share in total value of loans assumed to be a proportionate. Exclusions for low- or no-documentation loans and teaser-rate loans would have overlapped; taken together these borrowers accounted for about \$580bn of lending from 2003 to 2007.

loans with initial “teaser” rates, which were low introductory interest rates that would last for the first year or two of the mortgage (Mayer, Pence, and Sherlund 2009). By 2008, serious delinquency rates on these loans exceeded 30 percent. An affordability test that required mortgagors to demonstrate resilience to an interest rate stress up front would likely have curtailed non-prime lending on teaser rates significantly. If the test had prevented these loans altogether, it could have reduced mortgage lending by around \$370 billion (last line of Table 5).

Taken together, actions to restrict the borrowing of those that either could not certify their income or had stretched affordability characteristics would likely have materially dampened the surge in non-prime lending prior to the crisis. As Table 5 shows, we are left with a wide range of possible effects, but it seems plausible that combining loan-to-income and affordability rules would have moderated the scale of the household debt boom in the run-up to the crisis.

The macroeconomic benefit of any limits would incorporate the fact that they would have targeted the most highly indebted borrowers. Bunn and Rostom (2015), Andersen, Duus, and Jenson (2016), and Fagereng and Halvorsen (2016) show a correlation between pre-crisis household leverage and subsequent negative consumption responses. Although these studies do not demonstrate causality, they suggest that limiting leverage of the most highly indebted households could have a stronger aggregate effect on spending in a downturn than reducing debt uniformly across households. Targeted macroprudential interventions of this kind could therefore have been particularly effective in dampening the macroeconomic fallout from the crisis.

Table 6

Survey Data on Usage of Macroprudential Tools

| | Use of bank-focused tools <i>(positive countercyclical capital buffer, forward-looking provisions, caps on credit growth)</i> | Use of household-focused tools <i>(loan-to-income or debt-service-to-income limits)</i> | Both |
|--|---|---|------------------|
| All advanced economies (18) | 44% of countries | 33% of countries | 22% of countries |
| Advanced economies with financial stability committee with formal powers (5) | 60% of countries | 40% of countries | 40% of countries |
| Other advanced economies (13) | 38% of countries | 31% of countries | 15% of countries |

Note: We consider the 19 advanced economies covered in Edge and Liang (2017), minus South Korea, for which no data on tool usage is available. Numbers are based on country classification in Edge and Liang (2017) and survey responses on tool usage in IMF (2018a) that consider tools in use at the date of the survey. “Formal powers” refers to powers to act unilaterally or to issue “comply-or-explain” recommendations. Results for the United Kingdom have been adjusted to account for measures that had been agreed but were not yet binding at the date of the survey.

Could the Macroprudential Frameworks Set Up Since the Crisis Implement Such Policies?

Of the 58 countries surveyed in Edge and Liang (2017) that have created macroprudential frameworks since the crisis, 41 have set up multi-agency financial stability committees. Perhaps surprisingly, only 11 of these have formal powers, including either direct controls over macroprudential policy tools or the right to issue “comply or explain” recommendations to which other authorities are formally obliged to respond. The remaining cases rely on the voluntary cooperation of other regulators to achieve their policy aims.¹⁴

There is tentative evidence that financial stability committees with formal powers are more likely to act than government agencies exposed to short-term political pressures. Table 6, for instance, presents evidence from a sample of 18 advanced economies that 60 percent of countries with high-powered financial stability committees have taken bank-focused policy actions; this compares to 38 percent for countries where financial stability policy requires interagency cooperation. Similarly, countries with high-powered financial stability committees are also more likely to have used household-focused macroprudential tools. While the sample is small and causation is likely to run both ways, this evidence suggests that institutional frameworks do matter for mitigating biases towards inaction in the application of macroprudential policies.

If we line up the multi-agency financial stability committees in order of the powers at their direct disposal, two cases stand out at opposite ends of the spectrum. The UK Financial Policy Committee has a wide-ranging toolkit to achieve its

¹⁴Forbes (2018) provides a detailed summary of some of the macroprudential measures that different countries have put in place after the crisis.

mandate to protect and enhance the resilience of the UK financial system. At the other end of the spectrum, the US Financial Stability Oversight Council has few powers under its direct control. In this section, we compare and contrast these polar cases, and assess what they could do if faced with a rerun of the factors that led to the last financial crisis.

The US Financial Stability Oversight Council

The Financial Stability Oversight Council was set up in 2010 as part of the Dodd–Frank Act, and its ten voting members include the heads of existing regulatory agencies, including the Federal Reserve, Federal Deposit Insurance Corporation, Securities Exchange Commission, alongside the Secretary of the Treasury and one independent member with insurance expertise. It also has five non-voting members, including the Director of the Office of Financial Research, who serve in an advisory capacity. It is chaired by the Secretary of the Treasury, and most decisions are taken by majority rule.

Its overall mandate is to identify and respond to risks to US financial stability that could arise from the distress or failure of large, interconnected bank holding companies or nonbank financial companies. It is also charged with promoting market discipline by removing expectations that investors will be shielded from losses by the government. It is designed to facilitate information sharing between relevant regulatory agencies. Its only binding tool is the power to designate nonbank financial institutions deemed to be systemically important for enhanced supervision, a decision that has to be backed by two-thirds of the voting members. It has no other macroprudential powers. For other policy interventions, the Financial Stability Oversight Council can only issue recommendations to other regulators, not all of whom have an explicit financial stability objective (Kohn 2014).

We argue that the Financial Stability Oversight Council in its current form does not have sufficient powers to ensure financial stability in the face of a credit boom. The Council's track record to date supports this pessimistic assessment. First, consider the efforts to reform money market mutual funds. The runs on the Reserve Primary Fund in September 2008, followed by large outflows from money market mutual funds in general, revealed the fundamental vulnerability of these institutions to runs (Squam Lake Group 2011). Four years later in November 2012, the Financial Stability Oversight Council (unanimously) suggested three options for reform: making the shares in money market mutual funds have a floating rather than fixed value; mandating a 1 percent capital requirement along with a requirement that large withdrawals could be delayed; or mandating a 3 percent capital requirement that might be combined with other options. All of these actions would have reduced incentives to run a fund before it was no longer able to redeem shares at their face value.

Following heavy resistance, including criticisms that money market mutual funds were outside of the FSOC's remit, no final recommendation was issued (Cochran, Freeman, and Clark 2015). The Securities and Exchange Commission, as primary regulator for money market mutual funds, rejected the idea of capital requirements and ultimately passed a floating value option (on a 3 to 2 vote) that only applied to a subset of money market mutual funds. This took effect in October 2016.

A second case involves the long-running debate over attempts to reform US housing finance. There has been bipartisan political support for using government-sponsored enterprises like Fannie Mae and Freddie Mac to support the housing market, with some emphasis on making housing more affordable for lower-income borrowers (Rajan 2010). However, Alan Greenspan (2005), while Fed Chair and at the peak of his influence on public policy, repeatedly testified in favor of restraining the ability of government-sponsored enterprises to purchase private-label mortgage-backed securities on financial stability grounds, but to no avail. After the crisis, the Dodd–Frank Act did ban certain types of mortgages, such as interest-only mortgages or those with negative amortization. But it left the question of minimum down-payment restrictions to a group of six regulators involved in housing, which ultimately opted against introducing such a requirement. While risks in the housing market have significantly declined since the crisis, average loan-to-value ratios on mortgages are not lower than they were in the early 2000s. Furthermore, no US regulator has the ability to impose loan-to-income requirements, even if the Financial Stability Oversight Council wished to recommend this action.

Problems associated with lack of power of the Financial Stability Oversight Council would be mitigated if other authorities had the tools and incentives to act. For example, maintaining the stability of the financial system and containing the systemic risk that may arise in financial markets has long been central to the mission of the Federal Reserve (Kohn 2006). The Fed has a recognized set of powers including: conducting annual stress tests for large bank holding companies and the nonbank financial companies it supervises; setting countercyclical capital buffers for bank holding companies; imposing liquidity requirements on the largest and most complex financial institutions; and setting minimum margin requirements.

But the Fed’s powers are limited. For example, the Dodd–Frank Act curtailed the Fed’s ability to provide emergency lending to nonbanks. Moreover, the Fed lacks authority over many parts of the financial system and has no tools that can be used to tackle household debt vulnerabilities. A June 2015 “war game” exercise conducted by four Reserve Bank presidents concluded that the Fed had insufficient macroprudential powers to address a build-up in risks that resembled the earlier financial crisis (Adrian, de Fournouvelle, Yang, and Zlate 2017). Also, Fed officials have cast doubt on whether its mandate permits it to use monetary policy to act against a build-up in financial stability risks.

The UK Financial Policy Committee

The UK Financial Policy Committee was established in 2013. It has twelve voting members: the Governor of the Bank of England, four Deputy Governors, the Executive Director for Financial Stability, the head of the UK Financial Conduct Authority, and five independent external members. A representative of the Treasury attends the meetings as a non-voting member. Most of the Governors of the Bank of England also sit on the UK’s monetary policy and microprudential policy committees, which facilitates policy coordination. The large external membership, and the fact that decisions are taken primarily by consensus, means that external members have a strong voice.

The Financial Policy Committee is the most muscular macroprudential regulator in the world. It unilaterally sets the countercyclical capital buffer for all banks, building societies, and large investment firms operating in the United Kingdom, along with a countercyclical leverage buffer for large banks. Despite more than 70 countries reporting that they have put a countercyclical capital buffer framework in place, the UK Financial Policy Committee is one of only around 10 to have implemented it at positive rates to date and the only country to have released it in response to risks crystallizing. The Financial Policy Committee can also vary risk weights by sector for certain types of risk. It can impose limits on household borrowing via loan-to-income and loan-to-value restrictions. It advises on the scenarios used in the annual stress tests of the largest UK banking groups. It has successfully petitioned the government for additional powers. It can issue “comply or explain” recommendations to other regulators. In the past, it has made 18 recommendations to other regulators, all of which have been implemented. Finally, it makes an annual assessment on whether the perimeters of prudential regulation are drawn appropriately.

These powers are accompanied by a strong accountability framework. All members of the Financial Policy Committee are personally accountable to Parliament and typically provide testimony at least once a year. These testimonies follow the release of a biannual Financial Stability Report, which is increasingly designed to reach a wide audience to enhance public accountability.

If confronted by a rerun of the events leading to the financial crisis, the Financial Policy Committee would have the direct power to increase the resilience of the banking system by raising capital requirements via the countercyclical capital buffer rate, sectoral capital requirements, and countercyclical leverage buffers. While it does not have powers to direct changes in banks’ liquidity or funding requirements, it could issue comply-or-explain recommendations to the microprudential regulator to implement such changes. It seems plausible to us that, faced with evidence of mounting vulnerabilities throughout the early-to-mid 2000s, the Financial Policy Committee would have commissioned a stress test of the largest UK banking groups that would have assumed severe falls in house prices. Such an exercise might not have uncovered all the channels via which losses eventually transpired (for example, we doubt it would have been feasible to understand the full extent of losses that materialised on the seemingly very safe “AAA” tranches of collateralized debt obligations backed by mortgage securities¹⁵). Nevertheless, such tests would have exposed the fragile solvency and liquidity position of the largest banking groups (including their broker-dealer subsidiaries) at this point—thus signalling the need for higher capital and liquidity standards.

¹⁵That said, a paper written prior to the collapse in valuations of the senior AAA tranches of collateralized debt obligations argued that these assets were significantly overvalued given the likely states of the economy when defaults might occur—that is, given the systematic nature of the risks being borne by investors (Coval, Jurek, and Stafford 2009).

In addition, a regulator modeled on the Financial Policy Committee could have used its annual assessment to recommend changes to the regulatory perimeter to include other parts of the financial system like stand-alone broker-dealers that were not part of wider banking groups. Finally, the Financial Policy Committee could have guarded against vulnerabilities associated with household indebtedness by limiting the extension of certain mortgages with high loan-to-income ratios. Indeed, a loan-to-income limit and accompanying affordability test, similar to those considered above, was put in place by the Financial Policy Committee in 2014.

Conclusion

Could macroprudential policy frameworks have prevented the last crisis? Perhaps. There would have been challenges in spotting and responding to build-ups of risk in real time. But our analysis suggests that a macroprudential regime with a suitably strong mandate, coupled with powers to adjust financial system leverage and maturity/liquidity transformation and to limit household sector indebtedness, could have significantly ameliorated the macroeconomic fall-out from the collapse of the real estate bubble.

Are today's macroprudential regimes sufficiently well-equipped to do this? The US Financial Stability Oversight Council is not. The circumscribed structure of the Financial Stability Oversight Council reflects a political choice to limit the remit of financial regulation and, notably, to limit its ability to respond to financial sector developments outside the commercial banking system. While this may be deemed politically desirable, it would severely restrict the ability of US regulators to prevent a rerun of the crisis in the future. A macroprudential regulator modeled on the UK Financial Policy Committee would have the necessary mandate and powers, including in relation to household indebtedness. But a similar regime in a rerun of the crisis would still have required political backing to widen the perimeter of regulation to capture loosely regulated nonbank financial institutions and then to act aggressively.

This raises the important question of how much direct authority a macroprudential regulator requires. Many macroprudential regulators must rely on making nonbinding recommendations to other regulators. The evidence presented above suggests that one obvious risk of this arrangement is that the recipient of these recommendations does not share the macroprudential regulator's objectives, and no action will be taken. A less obvious risk is that a financial stability committee that lacks the authority to address risks will be tempted to see risks everywhere—after all, warning of such risks is costless and a useful way to hedge one's bets. For this reason, the warnings of a more powerful financial stability committee might be more targeted and informative.¹⁶

¹⁶ We would like to thank Sir Jon Cunliffe for suggesting this point to us.

A related question is how wide the remit of a macroprudential regulator should be. The evidence summarized above suggests that it may be necessary to take actions to ensure the resilience of both lender and borrower balance sheets, including by taking targeted action to prevent build-ups in household debt. While many countries have implemented policies aimed at preventing excessive levels of household debt, they remain controversial. Preventing a willing borrower and lender from consummating a mortgage contract where neither party is likely to default, for fear of the macroeconomic spillovers such contracts might create, is a more interventionist conception of macroprudential policy than one focused solely on resilience of the banking and financial system. It is not clear whether such interventions should be left to technical committees or to democratically elected governments (Balls, Howat, and Stansbury 2016; Tucker 2018).

A key challenge in making macroprudential policy effective is therefore to give the relevant financial stability committees clear powers and an appropriately wide remit, but also to put in place robust governance arrangements that ensure macroprudential policymakers are accountable for the way in which they use their tools.

■ *We thank Mathieu Chavez, Sir Jon Cunliffe, Beverly Hirtle, Don Kohn, Richard Sharp, Martin Taylor, Paul Tucker, Skander Van den Heuvel, and Andrei Zlate for helpful conversations about these issues. Kashyap's research has been supported by a grant from the Alfred P. Sloan Foundation to the Macro Financial Modeling (MFM) project at the University of Chicago and by the Chicago Booth Initiative on Global Markets and Fama-Miller Center. The views in this paper are those of the authors, and not necessarily those of the Bank of England or its policy committees.*

References

- Acharya, Viral V., Thomas Philippon, Matthew Richardson, and Nouriel Roubini. 2009. "The Financial Crisis of 2007–2009: Causes and Remedies." *Financial Markets, Institutions & Instruments* 18(2): 89–137.
- Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2016. "Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class." *Review of Financial Studies* 29(7): 1635–70.
- Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2017. "Dynamics of Housing Debt in the Recent Boom and Great Recession." *NBER Macroeconomics Annual* 2017, 32(1): 265–311.
- Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. 2016. "Vulnerable Growth." Federal Reserve Bank of New York Staff Report 794.
- Adrian, Tobias, Patrick de Fontnouvelle, Emily Yang, and Andrei Zlate. 2017. "Macroprudential Policy: A Case Study from a Tabletop Exercise." *Economic Policy Review* 23(1): 1–30.
- Adrian, Tobias, Michael Fleming, Or Shachar, and Erik Vogt. 2017. "Market Liquidity after the Financial Crisis." *Annual Review of Financial Economics* 9: 43–83.
- Aikman, David, Jonathan Bridges, Stephen Burgess, Richard Galletly, Iren Levina, Cian O'Neill, and Alexandra Varadi. 2018. "Measuring Risks to UK Financial Stability." Bank of England Staff Working Paper 738.
- Aikman, David, Jonathan Bridges, Anil Kashyap, and Caspar Siegart. 2018. "Would Macroprudential Regulation Have Prevented the Last Crisis?" Bank of England Staff Working Paper 747.
- Andersen, Asger Lau, Charlotte Duus, and Thais Lækholm Jensen. 2016. "Household Debt

- and Spending during the Financial Crisis: Evidence from Danish Mico Data." *European Economic Review* 89(C): 96–115.
- Avraham, Dafna, Patricia Selvaggi, and James Vickery.** 2012. "A Structural View of US Bank Holding Companies." Federal Reserve Bank of New York *Economic Policy Review* 7: 65–81.
- Bahaj, Saleem, Jonathan Bridges, Frederic Malherbe, and Cian O'Neill, C.** 2016. "What Determines How Banks Respond to Changes in Capital Requirements?" Bank of England Staff Working Paper 593.
- Balls, Ed, James Howat, and Anna Stansbury.** 2016. "Central Bank Independence Revisited: After the Financial Crisis, What Should a Model Central Bank Look Like?" M-RCBG Associate Working Paper 67.
- Bank of England.** 2017. *Financial Stability Report: June 2017*.
- Bassett, William F., Mary Beth Chosak, John C. Driscoll, and Egon Zakrajšek.** 2014. "Changes in Bank Lending Standards and the Macroeconomy." *Journal of Monetary Economics* 62: 23–40.
- Bernanke, Ben.** 2018. *The Real Effects of the Financial Crisis*. BPEA Conference Drafts, September 13–14.
- Borio, Claudio.** 2003. "Towards a Macroprudential Framework for Financial Supervision and Regulation?" *CESifo Economic Studies* 49(2): 181–215.
- Borio, Claudio, and William R. White.** 2004. "Whither Monetary and Financial Stability? The Implications of Evolving Policy Regimes." BIS Working Paper 147.
- Bridges, Jonathan, Chris Jackson, and Daisy McGregor.** 2017. "Down in the Slumps: The Role of Credit in Five Decades of Recessions." Bank of England Staff Working Paper 659.
- Brunnermeier, Markus K.** 2009. "Deciphering the Liquidity and Credit Crunch 2007–2008." *Journal of Economic Perspectives* 23(1): 77–100.
- Bunn, Philip, and May Rostom.** 2015. "Household Debt and Spending in the United Kingdom." Bank of England Staff Working Paper 554.
- Chavaz, Matthieu, and Andrew K. Rose.** Forthcoming. "Political Borders and Bank Lending in Post-Crisis America." *Review of Finance*.
- Chodorow-Reich, Gabriel.** 2014. "The Employment Effects of Credit Market Disruptions: Firm-Level Evidence from the 2008–9 Financial Crisis." *Quarterly Journal of Economics* 129(1): 1–59.
- Cochrane, Martha L., David F. Freeman, and Helen Mayer Clark.** 2015. "Money Market Fund Reform: SEC Rulemaking in the FSOC Era." *Columbia Business Law Review* 861.
- Coval, Joshua D., Jakub W. Jurek, and Erik Stafford.** 2009. "Economic Catastrophe Bonds." *American Economic Review* 99(3): 628–66.
- Duffie, Darrell.** 2018. *Prone to Fail: The Pre-Crisis Financial System*. Forthcoming.
- Edge, Rochelle M., and Nellie Liang.** 2017. "Who is in Charge of Financial Stability, Why, and What They Can Do." Hutchins Center Working Paper 32. <https://www.brookings.edu/research/who-is-in-charge-of-financial-stability-why-and-what-they-can-do/>.
- Eichner, Matthew J., Donald L. Kohn, and Michael Palumbo.** 2013. "Financial Statistics for the United States and the Crisis: What Did They Get Right, What Did They Miss, and How Could They Change?" In *Measuring Wealth and Financial Intermediation and Their Links to the Real Economy*, edited by Charles R. Hulten and Marshall B. Reinsdorf, 39–66. University of Chicago Press.
- Fagereng, Andreas, and Elin Halvorsen.** 2016. "Debt and Household Consumption Responses." Norges Bank Staff Memo No. 1.
- Financial Crisis Inquiry Commission.** 2011. *The Financial Crisis Inquiry Report: Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States*. Submitted pursuant to Public Law 111-21, January. Available at: <https://www.gpo.gov/fdsys/pkg/GPO-FCIC/content-detail.html>.
- Firestone, Simon, Amy Lorenc, and Benjamin Ranish.** 2017. "An Empirical Economic Assessment of the Costs and Benefits of Bank Capital in the US." Finance and Economics Discussion Series Working Paper no. 2017-034.
- Fleming, Michael J.** 2012. "Federal Reserve Liquidity Provision during the Financial Crisis of 2007–2009." *Annual Review of Financial Economics* 4: 161–77.
- Forbes, Kristin.** 2018. "Macroprudential Policy after the Crisis: Forging a Thor's Hammer for Financial Stability in Iceland." Report prepared at the request of the Task Force to review monetary and currency policies for Iceland.
- Gennaioli, Nicola, and Andrei Shleifer.** 2018. *A Crisis of Beliefs: Investor Psychology and Financial Fragility*. Princeton University Press.
- Gertler Mark, and Simon Gilchrist.** 2018. "What Happened: Financial Factors in the Great Recession." *Journal of Economic Perspectives* 32(3): 3–30.
- Greenlaw, David, Jan Hatzius, Anil K Kashyap, and Hyun Song Shin.** 2008. "Leveraged Losses: Lessons from the Mortgage Market Meltdown." In *Proceedings of the US Monetary Policy Forum*, pp. 7–59.
- Greenspan, Alan.** 2005. "Regulatory Reform of the Government-Sponsored Enterprises." Testimony before the Committee on Banking, Housing, and Urban Affairs, U.S. Senate. Speech 63, Board of Governors of the Federal Reserve System, US. April 6.

- Guerrieri, Luca, Matteo Iacoviello, Francisco Covas, John C. Driscoll, J., Michael T. Kiley, Mohammad Jahan-Parvar, Albert Queralto Olive, and Jae W. Sim.** 2015. "Macroeconomic Effects of Banking Sector Losses across Structural Models." Finance and Economics Discussion Series Working Paper 2015-044.
- Hanson, Samuel G., Anil K Kashyap, and Jeremy C. Stein.** 2011. "A Macroprudential Approach to Financial Regulation." *Journal of Economic Perspectives* 25(1): 3–28.
- Hirtle, Beverly.** 2016. "Bank Holding Company Dividends and Repurchases during the Financial Crisis." Federal Reserve Bank of New York Staff Report 666.
- Hirtle, Beverly, Anna Kovner, James Vickery, and Meru Bhanot.** 2016. "Assessing Financial Stability: The Capital and Loss Assessment under Stress Scenarios (CLASS) Model." *Journal of Banking & Finance* 69(Supplement 1): S35–S55.
- Holmquist, Elizabeth, and Josh Gallin.** 2014. "Repurchase Agreements in the Financial Accounts of the United States." FEDS Notes, Board of Governors of the Federal Reserve System, June 30.
- Homlström, Bengt, and Jean Tirole.** 1998. "Private and Public Supply of Liquidity." *Journal of Political Economy* 106(1): 1–40.
- International Monetary Fund (IMF).** 2005. *Global Financial Stability Report, April 2005: Market Developments and Issues.*
- International Monetary Fund (IMF).** 2018a. *The IMF's Annual Macroprudential Policy Survey—Objectives, Design, and Country Responses.*
- International Monetary Fund (IMF).** 2018b. *Global Financial Stability Report, April 2018: A Bumpy Road Ahead.*
- Jordà, Òscar, Moritz Schularick, and Alan M. Taylor.** 2013. "When Credit Bites Back." *Journal of Money, Credit and Banking* 45(S2): 3–28.
- Jordà, Òscar, Moritz Schularick, and Alan M. Taylor.** 2016. "The Great Mortgaging: Housing Finance, Crises and Business Cycles." *Economic Policy* 31(85): 107–152.
- Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig.** 2010. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans." *Quarterly Journal of Economics* 125(1): 307–62.
- Kohn, Donald L.** 2006. "The Evolving Nature of the Financial System: Financial Crises and the Role of the Central Bank." Speech given at the Conference on New Directions for Understanding Systemic Risk, New York, New York. May 18.
- Kohn, Donald L.** 2014. "Institutions for Macroprudential Regulation: The UK and the U.S." Speech at the Kennedy School of Government, Harvard University. April 17.
- Lallour, Antoine, and Hitoshi Mio.** 2016. "Do We Need a Stable Funding Ratio? Banks' Funding in the Global Financial Crisis." Bank of England Staff Working Paper 602.
- Martin, Christopher, Manju Puri, and Alexander Ufier.** 2018. "Deposit Inflows and Outflows in Failing Banks: The Role of Deposit Insurance." NBER Working Paper 24589.
- Mayer, Christopher, Karen Pence, and Shane M. Sherlund.** 2009. "The Rise in Mortgage Defaults." *Journal of Economic Perspectives* 23(1): 27–50.
- Mian, Atif, and Amir Sufi.** 2011. "House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis." *American Economic Review* 101(5): 2132–56.
- Mian, Atif, and Amir Sufi.** 2012. "What Explains High Unemployment? The Aggregate Demand Channel." NBER Working Paper 17830.
- Mian, Atif, Amir Sufi, and Emil Verner.** 2017. "Household Debt and Business Cycles Worldwide." *Quarterly Journal of Economics* 132(4): 1755–1817.
- Rajan, Raghuram G.** 2010. "Fault Lines: How Hidden Fractures Still Threaten the World Economy." Princeton University Press.
- Rosengren, Eric S.** 2014. "Broker-Dealer Finance and Financial Stability." Speech at the Conference on the Risks of Wholesale Funding in New York, New York. August 13, 2014.
- Squam Lake Group.** 2011. "Reforming Money Market Funds: A Proposal by the Squam Lake Group."
- Tarullo, Daniel K.** 2014. "Macroprudential Regulation." *Yale Journal on Regulation* 31(3): 505.
- Tucker, Paul.** 2018. *Unelected Power: The Quest for Legitimacy in Central Banking and the Regulatory State.* Princeton University Press.

The Value of US Government Data to US Business Decisions

Ellen Hughes-Cromwick and Julia Coronado

The US government is a major producer of economic and financial data, statistics, analysis, and forecasts that are gathered, compiled, and published as public goods for use by citizens, government agencies, researchers, nonprofits, and the business community. There is no market transaction in the publication and dissemination of these government data and therefore no market-determined value.

The purpose of this paper is to outline and augment our understanding of the value of government data for business decision-making. We provide an overview of the topic, including results from government reports and a private sector survey. We then provide concrete examples of how these government data are used to make business decisions focusing on three sectors: automotive, energy, and financial services. Examples of new initiatives by the federal government to open access to more data, exploiting technology advances associated with the internet, cloud storage, and software applications, are discussed. With the significant growth in the digital economy, we also include discussion and insights around how digital platform companies utilize government data in conjunction with their privately generated data (or “big data”) to foster more informed business decisions.

■ *Ellen Hughes-Cromwick is Associate Director and Senior Economist, University of Michigan Energy Institute, Ann Arbor, Michigan. She is also a Senior Advisor at MacroPolicy Perspectives LLC. Julia Coronado is President of MacroPolicy Perspectives LLC, New York City, New York. Hughes-Cromwick is the corresponding author at ehughesc@umich.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.33.1.131>

doi=10.1257/jep.33.1.131

Our exploration of the value of the public good provided by government data is necessarily qualitative, but a common theme is that for private firms, public data is an important complement and baseline to their own data. For example, in a 2017 panel discussion of the importance of government data, one participant noted that the big data now produced by many businesses are not sufficient to support optimal business decisions. Value is derived when “a firm’s own data are complemented with a wide range of data that are collected by the government. Federal data are comprehensive, covering the entire US, and, as a result, are useful for benchmarking and supplementing businesses’ own data. They’re also consistent with many data series spanning decades, allowing comparisons across place and over time” (Project at Brookings and American Enterprise Institute 2017; Brooks et al. 2017).

We believe that government support of data development and access to data is a competitive advantage for both existing and new US businesses. We see a risk that declining government support will lead to an erosion in the quality of public data and the value it provides to US businesses. As of FY 2017, total funding for the government’s 13 principal statistical agencies stood at \$2.257 billion (Office of Management and Budget 2018). By our calculations, this represents an 8.7 percent decline in real dollars from the 2004–2013 average budget for these agencies (based on Economics and Statistics Administration 2014, p. 13).¹ At a time when data capabilities and information technology are advancing rapidly, public data collection and dissemination requires ongoing investment and modernization to keep pace with rapid economic structural change.

Broad Assessments of the Value of Public Data to the Business Sector

Two US Department of Commerce reports and a recent survey of business economists provide broad-based assessments of the value of public data to the business sector.

US Department of Commerce Reports

A report from the Economic and Statistics Administration (2014) provides substantive documentation regarding the value of government data for professional managers at US businesses.

The report includes a summary of “government-data–intensive sectors” (GDIS) including businesses that “rely heavily on government data in their production processes” (p. 31). These include investment analysts, database aggregator firms, market researchers, benchmarkers, and others. The report estimated the 2012 GDIS

¹To estimate inflation-adjusted outlays, the authors used the Personal Consumption Expenditure price index, produced by the US Bureau of Economic Analysis. For comparability, costs of the 2010 Decennial Census are omitted from the 2004–2013 average, and preliminary costs of the 2020 Decennial Census are omitted from the government statistical budget in 2017.

revenues at \$220.8 billion (p. 41). This sector has grown substantially as digital platform companies combine government data with internally generated big data to create analytic tools and platforms that inform a host of business decisions.

A more recent comprehensive estimate of GDIS revenues will require an update of this data obtained from the 2012 Economic Census, and initial release of more recent data will begin in September 2019. However, other data sources strongly suggest that the GDIS sectors have been growing substantially since 2012. For example, the Census Bureau's 2015 report of "Statistics of US Businesses" (SUSB) includes information on payroll outlays and number of employees in GDIS. As of 2015, the payroll outlays were \$197.8 billion and 2.721 million employees, up 25.0 and 11.3 percent, respectively, since 2012.² The Bureau of Economic Analysis (BEA) data on GDP by industry includes gross output in current dollars for NAICS code 51930—internet publishing and broadcasting and web search. As of 2017, BEA estimates this industry had gross output of \$176.9 billion, a 92 percent increase since 2012, and well above the revised 2012 output of \$92.2 billion in output published in a November 2018 BEA release.³

Table 1 provides a snapshot on several private and public companies that rely substantially on government data to undertake their business activities in government-data-intensive sectors. Business revenues are substantial and have grown, in part because of new technologies enabling greater value creation through analytics. Growth in the value-added of government data has been enhanced by the ability to link directly to government data sources through application programming interfaces (API). Beyond this electronic access, businesses employ more sophisticated, cloud-based tools, which provide for the integration of government and big data to undertake analytics. Advancements in technology mean government data are now leveraged for even greater value across many different industries.

The Economic and Statistics Administration (2014) report also features a number of "data-driven" business decisions, which give concrete examples of the ways in which many firms use government data. For example, a large retailer used data from the American Community Survey (ACS) produced by the US Census Bureau to target customized inventories tailored to suburban and urban purchase attributes (p. 19). A small business in Texas received "customized market research from the US Commercial Service (in the International Trade Administration at the US Department of Commerce), which assisted the company in its penetration of

²The Statistics of U.S. Businesses (SUSB) dataset from the US Census Bureau can be found at <https://www.census.gov/programs-surveys/susb/data.html>. These totals include annual payroll outlays for NAICS codes 5191 (Other Information Services), 5313 (Other Activities Related to Real Estate), 5416 (Management, Scientific, and Technical Consulting Services), and 5419 (Other Professional, Scientific, and Technical Services). While not all of the activity in these NAICS codes can be attributed to support for data-driven business decisions, it does provide some sense of the magnitude of how government data generates value added in the business community.

³Bureau of Economic Analysis, https://apps.bea.gov/iTable/index_industry_gdpIndy.cfm. GDP-by-Industry, Underlying Detail of Industry, "Economic Accounts: GDP by Industry," "Table U: Gross Output by Industry," Billions of Dollars, November 2018.

Table 1
**Some Firms in the Government-Data-Intensive Sector:
 Revenue and Market Capitalization**

| <i>Company</i> | <i>Revenue (\$ millions)</i> | <i>Market cap (\$ millions)</i> |
|-----------------|------------------------------|---------------------------------|
| Public | | |
| Acxiom | 930 | 3,770 |
| IHS Markit | 3,890 | 20,160 |
| Nielsen | 6,660 | 9,350 |
| Redfin | 430 | 1,430 |
| Thomson Reuters | 11,410 | 31,600 |
| Zillow | 1,190 | 7,740 |
| Private | | |
| Bloomberg LP | 9,400 | NA |
| ESRI | 1,000 | NA |
| Haver Analytics | 3 | NA |
| Mapquest | 210 | NA |
| McKinsey | 10,000 | NA |
| Truven Health | 610 | NA |

Source: Data for public companies obtained from www.finance.yahoo.com; includes latest four-quarter trailing revenues and market cap as of October 11, 2018. Private company data are estimates from Gale Business Insights as of October 2018.

export markets (p. 21). Businesses use producer price data to inform price adjustments to sales and purchase contracts (p. 25). A large pet supplies retailer used Census Bureau data to optimize new store locations and to inform decisions about merchandise planning and advertising (p. 34).

Another way to gauge the use of government data by businesses is with Input-Output Accounts produced by the Bureau of Economic Analysis (2018). For input-output (IO) code 514, which includes data processing, internet publishing, and other information services, the BEA accounts indicate a total output of \$189 billion in 2016, up by 26.4 percent as compared to 2012.⁴ Admittedly, not all of the value of this industry is represented by government data. Even so, it represents only a portion of the government-data-intensive sector as defined in the Economic and Statistics Administration (2014) report.

The Economics and Statistics Administration (2015) followed up with a more focused study of business use in an assessment of the American Community Survey (ACS), an annual US Census Bureau survey of households that gathers detailed demographic data on jobs and occupations, educational attainment, home ownership, and other topics. As of November 2014, nearly 4,000 businesses subscribed to the ACS email updates, accounting for 12.3 percent of the subscriber base (p. 32).

⁴IO Code 514 includes NAICs codes 5182, 51911-2, 51919, and 51913. These industries are data processing, hosting & related services, libraries and archives, news syndicates, internet publishing & broadcasting & web search portals, and all other information services (NAICs codes can be found here: <https://www.census.gov/eos/www/naics/>).

The report gives a number of examples of how firms use the ACS. For example, businesses use it to inform their decisions about site selection and other commercial real estate decisions (pp. 33–34). Demographic information culled from the ACS is analyzed in conjunction with a businesses' proprietary information on sales in order to determine market share and other benchmark metrics (pp. 33–34). The ACS is used to develop business plans for product and marketing decisions. As part of this effort, businesses combine data on sales and store attributes with local demographic data to understand if they are positioning products properly to optimize sales (p. 34). This report highlights business demand for more and better government data to assist in their growth and development.

National Association for Business Economics Survey

The National Association for Business Economics (NABE) conducted a survey of its private sector members on the use of government data (for details, see Appendix 1). The survey was administered during April–May 2018 and included 14 questions regarding survey respondents' use of government data to inform business decisions. Just under 60 NABE members responded to the survey from a mixture of industries, including service industries like finance, insurance, and real estate, as well as goods-producing industries. Sixty-four percent of the respondents noted that their employer sells products and services through digital platforms.

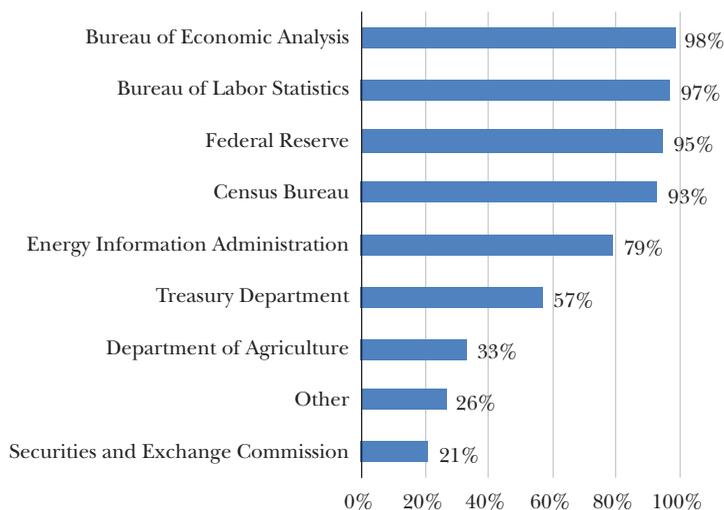
Ninety-five percent of the respondents replied yes to the question, "Are government data important to analyses and forecasting that drive business decisions?" Figure 1 displays responses to the survey question, "From which of the following agencies do you obtain data to inform business decisions at your firm, or firms with whom you work or consult?" When NABE members were asked how they rated the importance of specific types of government data they used from these agencies in order to inform business decisions, the top five responses were: 1) Employment and unemployment; 2) Prices and wages; 3) GDP; 4) Population; and 5) Income and profits.

Respondents were also asked to evaluate the usefulness of government data as inputs into a host of decision-making processes at their companies. On a scale of 1 to 5, with 1 being not critical and 5 being very critical, about half of respondents rated the following decisions as most critical (a response of 4 or 5): capital spending decisions; price-related decisions pertaining to cost-of-living adjustments for workers; finance-related decisions, such as discount rates for pension funds or recommendations regarding asset allocation; and interest rate decisions, such as when to borrow or lend and at what duration and/or cost.

Concrete examples of important uses of government data included the development of models used to project defense spending by industry, state, and occupation; infrastructure investment spending; or industry footprint analysis at the state and regional level. One respondent noted that government data on health care allows for the development of models to help healthcare facilities decide which services to expand geographically and how many providers and support staff would be required to meet projected demand. Another respondent noted that the firm

Figure 1

NABE Survey Question: “From which of the following agencies do you obtain data to inform business decisions at your firm, or firms with whom you work or consult?”



Source: National Association for Business Economics (NABE) Survey on Use of Government Data to Drive Business Decisions, April–May 2018.

uses data from the US Department of Agriculture on production and prices to estimate demand for diesel engines in the agriculture sector. Respondents in the finance, insurance, and real estate services industries are particularly intensive users of government data. Another respondent noted the importance of data use for “[s]etting loan and deposit rates. Keeping senior management informed of key government data releases and implications for financial markets.”

There is likely to be sample selection bias associated with these survey results—after all, those that value government data are most likely to respond to the survey—but the types of data valued and examples of use nonetheless provide insight into how the public good of government data enhances business decision-making.

A more recent survey conducted by Bi-Survey.com (2018) captured the views of over 600 respondents regarding the type of data used for decision-making. Despite the growth in large, internally generated datasets, this survey found that the growth in the use of external data sources for business decision-making was somewhat higher. Over one-half of the companies surveyed use at least five external data sources, while nearly 25 percent stated that they use more than 10 external data sources. The survey did not ask for the source of the external data, and we are not able to ascertain if most, if not all, of these external data sources are published by government entities. However, the results confirm that external data, including public data, are a complement to the increased generation of large amounts of internal data that companies produce. Growth in the use of big data suggest increased value created by public data.

Examples from Three Industries: Automotive, Energy, and Financial Services

Automotive Sector

The US automotive industry, and more broadly, the transportation sector of the economy, is large and diverse. In 2017, the value-added of automotive transportation-related industries was 3.7 percent of GDP, or \$713.5 billion.⁵ Consumer and business purchases of vehicles exhibit procyclical behavior. Both the supply chain and product sales of the industry are global in nature. Thus, business decisions rely on insights and forecasts regarding economic activity, including short- and long-term behavior of GDP, inflation, interest rates, commodities, and exchange rates—for the US economy and global economies.

Table 2 shows examples of the data required to inform business decisions in the automotive industry in the short-run and the long run (based on the career experience and ongoing professional contacts of the authors). In addition, Manyika et al. (2013) describe government data used in the transportation services industry to drive business decisions. Transportation sectors included in their study are marine shipping, air, passenger autos, and rail.

An example of a short-run business decision is the modeling of automotive demand conditions and the near-term outlook, which is necessary to make informed decisions about production rates at assembly plants. US government data are combined with internally generated and other private sector data, allowing experts in each of these subject areas to collaborate with team members from other functions within the company: for example, marketing and sales, finance, credit, product development, and the business operations running the plants. In turn, the US government data support sales forecasts, management of desired inventory levels, and expected competitor behavior and pricing in order to make business decisions regarding production at the assembly plant level.

Long-term business decisions at an automotive company require additional data and modeling to perform analyses, including investment decisions regarding assembly plant expansion or site location, and assumptions of revenue growth based on vehicle industry pricing projections. Businesses also rely on government data for their financial forecasts which, in turn, influence pricing of leases and loans, healthcare cost projections, cash management, pension funding, securitization funding, and other financial decisions. As another example, the use of government data emanating from GPS satellites has improved decisions around supply-chain management, logistics, mapping, and route planning (Manyika et al. 2013, p. 31).

The ongoing evolution of the transportation industry into electrified, connected and automated vehicles (EVs and CAVs) rely on government data as well. The Center for Open Data Enterprise (2017) summarized the results of a White House Roundtable on Open Data for Economic Growth, held on July 25, 2017.

⁵Bureau of Economic Analysis, GDP by Industry data as of November 1, 2018.

Table 2

US Government Data Used for Short-Term and Long-Term Auto Industry Decisions

| <i>Short-run indicators</i> | <i>Long-run indicators</i> |
|---|---|
| Auto sales (BEA) | Auto sales (BEA) |
| Consumer credit (Federal Reserve) | Auto production and assemblies (Federal Reserve) |
| Consumer price index for new vehicles (BLS) | Consumer credit (Federal Reserve) |
| Consumer price index for all items (BLS) | Consumer price index for new and used vehicles (BLS) |
| Disposable personal income (BEA) | Consumer price index for all items (BLS) |
| Employment and Unemployment (BLS) | Disposable personal income (BEA) |
| Energy prices (BLS and EIA) | Energy prices (BLS and EIA) |
| GDP (BEA) | GDP (BEA) |
| Interest rates (Federal Reserve) | Consumer spending and income distribution (BLS and Census) |
| Inventories (Census) | Household wealth (Federal Reserve) |
| Regional income, prices, and consumer spending (BEA and Census) | Industrial production and capacity utilization (Federal Reserve) |
| | Interest rates (Federal Reserve) |
| | International trade and investment (BEA and Census) |
| | Inventories (Census) |
| | Population (Census) |
| | Regional and state GDP, disposable personal income, and prices (BEA, BLS, Census) |
| | US Federal Budget (OMB, CBO, Treasury) |
| | US Federal Government Debt (Treasury) |
| | Vehicle miles traveled and travel attributes (DOT) |

Source: Author's assessments based on professional work at Ford Motor Company.

Note: BLS is Bureau of Labor Statistics; BEA is Bureau of Economic Analysis; EIA is Energy Information Administration; OMB is Office of Management and Budget; CBO is Congressional Budget Office; DOT is Department of Transportation.

The transportation sector participants included auto manufacturers, auto insurance companies, public transportation organizations, and companies that provide vehicle-sharing and other innovative models (p. 5). Participants noted that they use information on public transit systems to develop web and mobile applications for consumers, while auto companies use government data on transportation to support the development of autonomous vehicles. Insurance companies rely on transportation data, specifically accident statistics to estimate incidence probabilities in order to derive market pricing for premiums.

Energy Sector

Firms in the energy sector include crude oil producers, refiners, oil servicing companies, electric utilities, natural gas producers, coal companies, nuclear companies, pipeline producers, and suppliers of energy-related equipment and

components such as windmill turbines, solar panels, other renewable energy sources, and battery storage units.

The Energy Information Administration (EIA) is an independent statistics and analysis agency within the US Department of Energy, created in 1977 in the aftermath of the first OPEC oil shock (Government Printing Office 1977). It provides timely energy statistics and forecasts on every dimension of the energy sector. These data include sources and uses of energy by type and geography, prices of energy by type, short- and long-term forecasts of the energy sector, including several types of disaggregation (for example, by country, by region within the United States, and by end-use such as residential, commercial building, manufacturing, and transportation). The EIA also measures energy imports and exports and provides data on drilling activity in the offshore and shale fields across the United States. By law, EIA's data, analyses, and forecasts are independent of approval by any other officer or employee of the US government.

For energy-intensive industries such as durable goods manufacturing, chemicals, construction, and transportation, obtaining information about energy price trends is vital to gauging the outlook for energy costs, demand, and supply conditions. Energy consulting companies depend on government data as a starting point for market analysis. For example, an analyst of these data at an energy consulting firm noted that weekly data on pricing and inventories drive the short-term price of crude oil and energy-related financial products in the futures markets. In turn, energy producers utilize the futures markets in order to hedge against adverse swings in pricing and to inform decisions about production rates. Decisions on refinery runs rely on inventory and demand information as well as product pricing. Import and export decisions are based on whether the crude and crude-related products will be needed in the US market, which starts with understanding recent data and trends. Investments about physical storage are based on these data. One concrete example is the use of EIA diesel fuel price data in rate-setting for interstate trucking. EIA energy consumption surveys for building infrastructure are used as benchmarks for many private decisions on utility services and design criteria for offices, schools, hospitals, shopping malls, and private residences.

Because of the comprehensive nature of the statistics and forecasts from the Energy Information Administration, very few business decisions in the energy sector are not informed by these government data. Private companies lack the legal authority of the EIA to acquire and disseminate data, and so are unable to duplicate EIA's breadth and depth of transparency. It is, arguably, one of the most valued government datasets available to the public for free.

For career development purposes, Carnegie Mellon University (2018) posts a comprehensive list of 30 energy consulting firms on its website. All of these companies would, in principle, rely on government data in order to undertake analysis and recommendations for their clients. Government data from the Energy Information Administration is vast and free.

Beyond the energy sector, market pricing of the outlook for energy prices affects the outlook for inflation, interest rates, and a wide range of asset prices,

including the value of the dollar. Energy futures prices are also a function of Energy Information Administration data on demand and supply statistics. These datasets also underpin assessments of inflation which are embedded in the prices of Treasury inflation-protected securities (TIPS). These, in turn, influence economists' and policymakers' forecasts of inflation, which are a key input for monetary policy decisions and expectations for consumer spending.

Financial Services Sector

Financial services firms include commercial banks, asset management firms, equity brokerages, credit unions, and finance companies. Financial services firms are arguably one of the most intense users of US government data, employing data from the Securities and Exchange Commission (SEC), finance-related data from the US Department of the Treasury, and all types of economic and demographic data.

A central application of government data is the stress-testing of the balance sheets of "systemically important financial institutions." This exercise, which must be completed at least annually (large banks have to test themselves semiannually) requires firms to estimate the impact on their capital bases of two adverse economic and financial scenarios (Board of Governors of the Federal Reserve System 2018).

The stress-testing process involves extensive econometric modeling. Credit losses are projected using a combination of borrower characteristics and macroeconomic variables. Prominent in the latter category are GDP and its components, unemployment measures, and personal income. Both state-level and US aggregate data are utilized. Considerable use is made of Federal Reserve data: household debt and asset levels from the Financial Accounts, as well as data on interest rates, consumer credit, bank assets and liabilities, and money supply. When banks monitor credit conditions for the credit management process, they use aggregate data on corporate and household debt, along with the Federal Reserve's survey of lending conditions called the Senior Loan Officer Survey. National and regional house price data are produced by the Federal Housing Finance Agency. Strategic planning exercises also often employ US Census Bureau data containing a wide range of information on demographics and wealth.

Financial companies described a number of uses of "open" (including US government) data at a 2017 forum held by the Center for Open Data Enterprise (2017, pp. 3–4), in collaboration with the Office of Management and Budget. For example, open data helps financial firms assess businesses seeking financing, from startups and small businesses, to larger companies. It informs a variety of investment decisions involving companies, specific sectors, real estate, currencies, commodities, and other assets. Financial firms use data on companies, professional licenses, property, court records, and more to detect fraud and mitigate risks. Open data aids firms in their research on national and global financial outlooks to understand consumer behavior, identify and quantify risks, and optimize their strategies. Financial institutions use demographic and social, economic, and labor data to assess loan applicants, which allows for the possibility of offering loans to those with no or limited credit history.

Data on consumer credit quality and economic conditions drive decisions and credit allocation in the primary and secondary markets for mortgage finance.

Financial firms often forecast aspects of the US economy, both for internal planning and as a product used by clients. The forecasts rely on a broad range of data on output, prices, labor markets, income, and other areas. Macroeconomic forecasting models rely heavily on public data. Macroeconomic Advisers' US model, for example, uses over 1,000 variables, with over 90 percent represented by government data (according to Ben Herzon, an Executive Director at the firm).

Corporate executives in the financial services industry recognize the critical role that government data perform in their decision-making. The chief executive officer of Northern Trust, Michael O'Grady, stated:

Northern Trust, and the financial services sector more generally, relies heavily on the economic data produced by governmental agencies. Stress testing for capital adequacy, performance planning, and risk measurement would be nearly impossible without it. Timely and high-quality information not only aids our operations, it contributes to the stability of our financial system. Our industry strongly supports adequate investment in this very critical resource.⁶

Open Data Initiatives and Value Estimation

Most government data are publicly available, free of charge, and available for download from government websites. However, the federal government has been undertaking many activities to provide open access to more data by utilizing technology advances that in turn facilitate the development of new business models, software applications, and strategic uses of public data. Open format data are machine readable through "application programming interfaces" (APIs) at zero cost and are available to any sector of the economy—government, academia, business, and other organizations (Dietrich et al., *Open Data Handbook*, undated).

Open government data is recognized by many government officials and business people as a strategic asset for economic growth and business opportunity in a world where investment is increasingly driven by intellectual capital (Zinnbauer 2018). The World Bank (2014) documented case studies of companies that have grown to valuations of more than \$1 billion through the use of open data, including the real estate company Zillow and the navigation service Waze, among many others.

The McKinsey Global Institute (MGI) published a report that estimated the value of open data in seven sectors: education, transportation, consumer products, electricity, oil and gas, health care, and consumer finance (Manyika et al. 2013, p. 9). The estimate of value added from open data was \$300–450 billion in health care, and \$210–\$280 billion in consumer finance. For the electric power sector, Manyika et

⁶Statement provided to the author by Carl Tannenbaum, Executive Vice President and Chief Economist at Northern Trust.

al. (p. 60) highlighted the importance of data from government regulatory agencies and from the Energy Information Administration as inputs for capital investment and productivity performance. The Federal Energy Regulatory Commission (FERC) is responsible for regulating the interstate transmission of natural gas, oil, and electricity. It also regulates natural gas and hydropower projects and does environmental impact assessments and cost calculations, which inform the assessment of pass-through of capital investment outlays to consumers in the form of rate increases.

One potential value of government data results from increased competition and efficiency. This possibility is acknowledged in the disclosures of large, global information providers that provide research, analytics, and forecasts to a substantial business client base. In their 2017 annual report, the firm IHS Markit (2017) alluded to the importance of government data in their “Risk Factors” description:

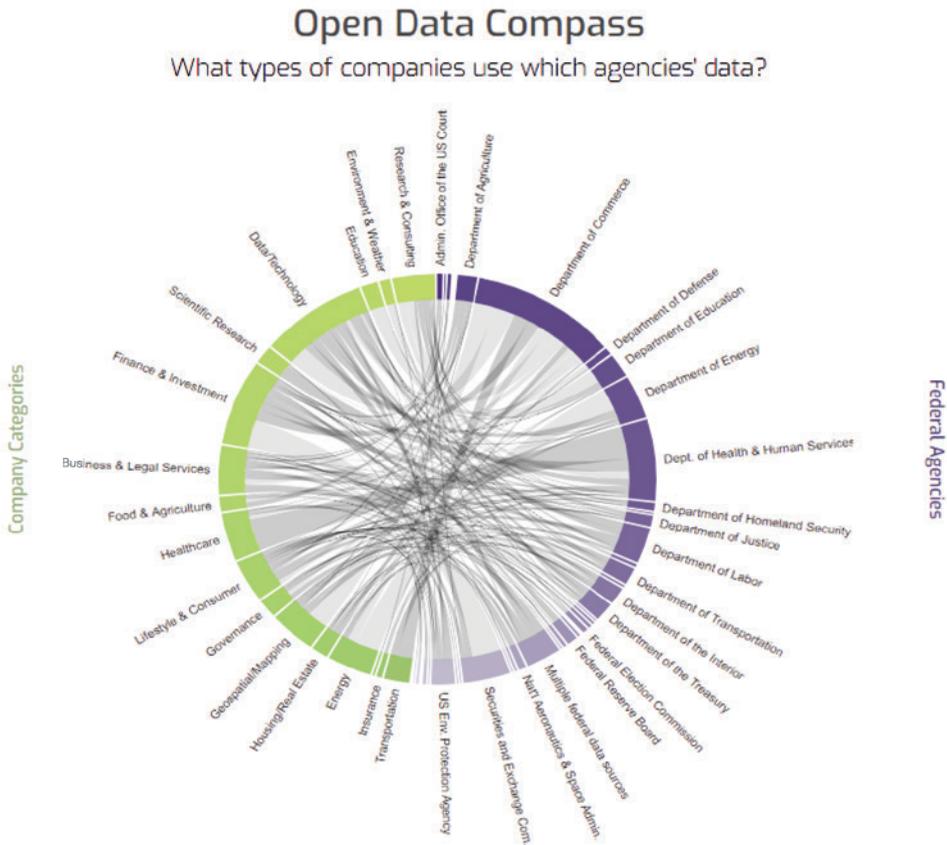
Some of the critical information we use in our offerings is publicly available in raw form at little or no cost. The internet, widespread availability of sophisticated search engines, pervasive wireless data delivery and public sources of free or relatively inexpensive information and solutions have simplified the process of locating, gathering, and disseminating data, potentially diminishing the perceived value of our offerings. While we believe our offerings are distinguished by such factors as currency, accuracy and completeness, and our analysis and other benefits, our customers could choose to obtain the information and solutions they need from public, regulatory, governmental or other sources. To the extent that customers become more self-sufficient, demand for our offerings may be reduced, and our business, financial condition, and results of operations could be adversely affected (p. 21).

Magalhaes and Roseira (2017) undertook an innovative approach to assess the use of open government data at 178 firms across nine US industries. They analyzed 178 companies that use open government data, which they then classify into twelve categories depending on how the data are used to create value. Their results show that data can be leveraged as business intelligence, process optimization, product/service improvement, and research and development (pp. 7–8).

The Center for Open Data Enterprise (2017) has also spurred other research on how government data are used by businesses. It developed the “Open Data Impact Map” which includes more than 500 examples of US firms that use open government data as a business resource (p. 2).

New York University hosts the Governance Lab, or GovLab, which “works to improve people’s lives by changing how we govern, using technology-enabled solutions and a collaborative, networked approach.” GovLab initiated Open Data 500, an effort to focus on identifying companies using government data for a variety of business objectives (<http://www.opendata500.com/us/>). Open Data 500 compiled the flows from government data agencies to sectors of the US economy. Figure 2, taken from their website, shows sources of government data by sector of the economy. Finance and technology sectors are the largest recipients of government

Figure 2
Sources of Government Data by Sector of Economy



Source: The GovLab @ NYU Open Data. 500 <http://www.opendata500.com/us/>.

data inflows. Open Data 500 also focused on small and mid-sized businesses, many of which were startups that relied heavily on open data for their business success.

A case study on the Kellogg Company published by the Center for Open Data Enterprise and Accenture (2017) analyzed how government data helps “guide business investments, develop new products and services, and foster innovation” (p. 4). The company employed data using “problem-centric” and “discovery-centric” approaches (p. 4). A problem-centric approach combines proprietary and publicly available data to address specific organizational challenges. A discovery-centric approach allows a “data analytics team to identify new correlations and trends by fostering intellectual curiosity” (p. 4). Both approaches help the company generate increased revenue, reduce costs, and better meet customer needs.

One database of substantial benefit to the private sector has been the open access to National Oceanic and Atmospheric Administration weather data. It facilitated the creation of The Weather Company, a business whose digital assets were

acquired by IBM for \$2 billion in 2015 (as reported in McMillan 2015). The US Department of Commerce entered into an agreement with the Weather Company and other business partners to unleash even more weather-related big data for business use. IBM subsequently purchased The Weather Company and has expanded the application of weather data to business decisionmaking with new artificial intelligence and modeling tools (IBM undated). These data drive business decisions about insurance pricing, and retail and healthcare sector preparedness related to storm activity, to name just a few uses. Experts at the Center for Open Data Enterprise are expanding the documentation and analysis of open data benefits to private businesses (Gurin 2017).

Use of Government Data by Digital Platform Companies

Digital platform companies, which rely on web-based transactions and information exchange with customers and other businesses, often combine their organically grown data with open government data to produce analytics that contribute to increased operational efficiencies, reduction in costs, and inventory and distribution management, and offer new revenue-generating services. We interviewed three data scientists and economists at digital platform companies who asked that their comments remain anonymous. All three indicated they use government data from different agencies to supplement the big data derived from their online businesses.

One company has an application that uses government data on retail market activity and interest rates to engage with their online customers. This application is viewed as a marketing tool by management and has been determined to be effective in engaging with customers and providing them with valuable information. Extensive use of data from the American Community Survey, the Consumer Expenditure Survey, Treasury Department data on aggregate tax receipts, Census Bureau data on demographics and retail sales, Department of Commerce data on weather, and labor market data were cited as important sources. The company uses these data to gauge emerging market trends, as well as the market size, which serves as a benchmark against which the digital platform company can assess its competitiveness and market growth.

Hiring decisions are often especially reliant on analysis of government data. To determine the need for employees in certain markets, some companies used Census Bureau, Bureau of Economic Analysis, and Bureau of Labor Statistics data to quantify the market size for their products and services. An analysis of state and metro area data and immigration and demographic trends allowed company analysts to make recommendations to management regarding the size of the professional labor force to hire in order to meet projected demand. Since labor costs are a substantial factor input for this firm, analysts were able to conduct a “bottom up” approach in order to match labor with demand conditions in local markets.

Conclusion

The value of government data is difficult to measure, but it is clearly a substantial strategic asset for the US business sector. Such data are used by a wide range of companies from auto producers to digital platform companies, and for purposes that include production and investment decisions, marketing and inventory management, and long-range strategic planning. They are also of paramount importance in undertaking assessments of the financial health of banks and the overall stability of the US financial system.

The value of government data seems to be increasing. Technology advances have allowed for an ever-increasing amount of data to be made easily accessible to citizens, businesses, and state and local governments. Data have growing added value for businesses in an information economy increasingly driven by intellectual capital, and this has allowed for new business formation and facilitated transparency and competition. While companies are generating ever-increasing amounts of big data from their own operations, it is often the combination of proprietary data with comprehensive government data that provide critical context and allow for maximum strategic benefit (a public good externality). Given the relatively low cost of government data production and the still rapid pace of information-related technological change, it is sound public investment to support budgets that will allow the US government statistical agencies to expand the quality, scope, access, and timeliness of their efforts.

■ *We are grateful to the staff at the National Association for Business Economics, including Tom Beers, Chris Jonas, and Michele Soresi for their help in conducting the survey. Many thanks to economists at over 20 companies who provided information and insights regarding the topic. We would also like to thank Maurine Haver for her deep knowledge of government data and advice on the topic. Special thanks also go to Carl Tannenbaum, Executive Vice President and Chief Economist for Northern Trust, who provided insights about government data usage in the financial services sector.*

References

Bi-Survey.com. 2018. "How Many Data Sources Do Companies Rely on for Decisionmaking?" <https://bi-survey.com/data-sources>.

Board of Governors of the Federal Reserve System. 2018. "Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule—February 2018." <https://www.federalreserve.gov/>

[publications/2018-february-supervisory-scenarios-for-annual-stress-tests.htm](https://www.federalreserve.gov/publications/2018-february-supervisory-scenarios-for-annual-stress-tests.htm).

Brooks, Authur, Robert Rubin, Dian Whitmorer, Rebecca Blank, Ellen Davis, Martin Feldstein, Torsten Slok, and David Leonhardt. 2017. *The Vital Role of Government Statistics: Strengthening Research, Governance, and Innovation*. Transcript from Panel Discussion, AEI Auditorium, Washington, DC,

March 2. American Enterprise Institute and the Hamilton Project at Brookings. <https://www.aei.org/wp-content/uploads/2017/02/170302-AEI-Vital-Role-of-Government-Statistics.pdf>.

Bureau of Economic Analysis. 2018. *Input-Output Accounts Data*. <https://www.bea.gov/industry/input-output-accounts-data>.

Carnegie Mellon University. 2018. “Energy Consulting Firms.” <https://www.cmu.edu/career/documents/college-specific-resources/coe-grad-undergrad/company-guides/energy-consulting-firms.pdf>.

Center for Open Data Enterprise. 2017. “2017 Open Data Roundtable: Roundtable on Open Data for Economic Growth.” <http://reports.opendataenterprise.org/2017OpenDataRT1-EconomicGrowth.pdf>.

Center for Open Data Enterprise and Accenture. 2017. “Government Data for Business Innovation in the 21st Century.” https://www.accenture.com/t20180103T192103Z__w__/_us-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Dualpub_26/Accenture-Open-Data-Executive-Summary.pdf.

Dietrich, Daniel, Jonathan Gray, Tim McNamara, Antti Poikola, Rufus Pollock, Julian Tait, and Ton Zijlstra. No date. *Open Data Handbook*. <http://opendatahandbook.org/guide/en/what-is-open-data/>.

Economics and Statistics Administration, US Department of Commerce. 2014. *Fostering Innovation, Creating Jobs, Driving Better Decisions: The Value of Government Data*. July. <https://esa.gov/sites/default/files/revisedfosteringinnovationcreatingjobsdrivingbetterdecisions-thevalueofgovernmentdata.pdf>.

Economics and Statistics Administration, US Department of Commerce. 2015. *The Value of the American Community Survey: Smart Government, Competitive Businesses, and Informed Citizens*. April. <https://www.esa.gov/sites/default/files/the-value-of-the-acs.pdf>.

Government Printing Office. 1977. Department of Energy Organization Act. <https://www.gpo.gov/fdsys/pkg/STATUTE-91/pdf/STATUTE-91-Pg565.pdf>.

GovLab @ NYU Open Data. Undated. “Open Data Compass.” <http://www.opendata500.com/us/>.

Gurin, Joel. 2017b. “How Open Data Fuels Key Business Sectors.” *fedscoop*, August 16. [https://](https://www.fedscoop.com/open-data-fuels-key-business-sectors/)

www.fedscoop.com/open-data-fuels-key-business-sectors/.

Hamilton Project of Brookings and American Enterprise Institute. 2017. “The Vital Role of Government Statistics in Strengthening Research, Governance, and Innovation.” Panel of Experts on March 2, 2017. http://www.hamiltonproject.org/events/the_vital_role_of_government_statistics_strengthening_research_governance_a.

IBM. Undated. “IBM and the Weather Company Deliver New AI-Driven Decision-making Tools.” <https://newsroom.ibm.com/the-weather-company?item=30634>.

IHS Markit. 2017. Annual Report 2017. Available at: <http://investor.ihsmarkit.com/phoenix.zhtml?c=188457&p=irol-financial>.

Magalhaes, Gustavo, and Catarina Roseira. 2017. “Open Government Data and the Private Sector: An Empirical View on Business Models and Value Creation.” *Government Information Quarterly*. <http://dx.doi.org/10.1016/j.giq.2017.08.004>.

Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve Van Kuiken, and Elizabeth Almasi Doshi. 2013. *Open Data: Unlocking Innovation and Performance with Liquid Information*. McKinsey Global Institute. <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>.

McMillan, Robert. 2015. “IBM Buys Weather Company’s Digital Assets, Expanding Move into Data Crunching.” *Wall Street Journal*, October 28.

Office of Management and Budget, Executive Office of the President. 2018. *Statistical Programs of the United States Government*. <https://www.whitehouse.gov/wp-content/uploads/2018/05/statistical-programs-2018.pdf>.

US Census Bureau. 2015. *2015 SUSB Annual Data Tables by Enterprise Industry*. <https://www.census.gov/data/tables/2015/econ/susb/2015-susb-enterprise.html>.

World Bank. 2014. “Open Data for Economic Growth.” June 25. Transport & ICT Global Practice. <http://documents.worldbank.org/curated/en/131621468154792082/pdf/896060REVISED000for0Economic0Growth.pdf>.

Zinnbauer, Dieter. 2018. “Open Government—Open for Business?” Open Data Partnership. https://www.opengovpartnership.org/sites/default/files/OGP_Open-for-Business_20180515.pdf.

On the Controversies Behind the Origins of the Federal Economic Statistics

Hugh Rockoff

Our federal economic statistics originated in the economic and political divisions in the United States and the bitter debates over economic policy they engendered at the end of the 19th century and during the world wars and Great Depression. Workers were angry because they believed that they were being exploited by robber barons who were capturing all of the benefits of economic growth, while employers were just as sure that the second industrial revolution had brought workers an unparalleled increase in real wages. Other debates centered on the effects of unrestricted immigration on wages and employment opportunities of native-born Americans, on the effects of tariffs on prices paid by consumers, on the effects of frequent financial panics on employment, and, during the world wars, on the effects of wage and price controls on the living standards of workers. Participants on all sides of these debates believed that nonpolitical and accurate statistics constructed by experts would help to win support for the policies they favored.

In most cases, the development of these statistics was led by individuals, private organizations, and state governments, although the federal government eventually took over the role of producing these statistics on a regular basis. Here I provide brief histories of the origins of US statistics on prices, national income and product, and unemployment to illustrate this story.

■ *Hugh Rockoff is Distinguished Professor of Economics, Rutgers University, New Brunswick, New Jersey, and a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His email address is rockoff@econ.rutgers.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at

<https://doi.org/10.1257/jep.33.1.147>

doi=10.1257/jep.33.1.147

Price Indexes

The honor of creating the first price index probably belongs to the Italian G. R. Carli, who published an index based on the prices of grain, wine, and oil in 1764 as part of an examination of the effects of precious metals from the New World on the European price level (Mitchell 1938, p. 7). The origins of the US government price indexes can be traced at least to the 1860s. During the Civil War, Secretary of the Treasury Salmon Chase published averages of prices, perhaps with the intention of showing that inflation—prices about doubled in the North during the war—was not as extreme as some critics of the government’s financial policies suggested.

For economists, the work of William Stanley Jevons (1863) has long been considered transformative because it marked a sharp break between statisticians as compilers of statistics and economists as producers and users (S. Stigler 1982). Jevons did not weight the individual prices he collected. Credit for weighted indexes belongs, as much as any economic ideas have identifiable inventors, to Etienne Laspeyres and Hermann Paasche (Persky 1998). Jevons computed price relatives (ratios of the price in a given year to the price in the base year) and then took a geometric average. A geometric average avoided distortions resulting from very large increases in individual prices. Jevons, moreover, went on to use the quantity theory of money to explain the inflation. He found that prices had risen about 30 percent between 1848 and 1860 and attributed this to the discoveries of gold in the mid-1800s. He confirmed his explanation by estimating the amount of gold in use, and showing that it had increased by about the same percentage as prices. He then explored the effects of the inflation on various classes such as bondholders, wage earners, and so on. Jevons’s choice of the geometric mean motivated the important work of British economist and statistician Francis Y. Edgeworth (1887, 1889). Edgeworth showed that the best mean depended on the distribution of prices, and that the margin of error associated with a particular mean depended on how the actual distribution differed from what was assumed (S. Stigler 1978, p. 297–98).

A crucial step in the path to our modern price statistics was the establishment of the US Bureau of Labor in 1884.¹ Organized labor wanted an organization that would fight for the rights of labor, while employers resisted. The compromise was an agency that would simply collect and distribute statistics about labor. President Chester Arthur chose Carroll D. Wright, who had built a reputation for competence and integrity as head of the Massachusetts Bureau of the Statistics of Labor—the first such agency—and who was acceptable to both labor and capital, to direct the new federal agency.

In the early 1890s, the Bureau of Labor produced indexes of wages and prices at the behest of the Aldrich Committee—named after Senator Nelson Aldrich, a Republican who later played an important role in establishing the Federal Reserve. The underlying motivation for the study, I surmise, was labor’s claim that wages

¹Weiss (1955) provides a concise but informative history of the price indexes produced by the Bureau of Labor Statistics. Duncan and Shelton (1978) provide a chronology of the work of the Bureau to 1976.

had fallen while prices had risen as a result of Republican tariff increases; a claim that some committee members, including Aldrich, believed could be laid to rest with comprehensive statistics of untainted provenance. The Committee's report was based on data collected by a research group directed by Roland Falkner, a professor of statistics at the University of Pennsylvania.² The group painstakingly collected 52,393 price observations (Conforti 2016, p. 3)—the era of big data had arrived! With this data in hand, Falkner produced wholesale and retail price indexes stretching from 1860 to 1891. A survey of expenditures by “normal” households—no more than five children, did not own their own home, among other characteristics—served as the basis for weighting the prices. Data on the number of workers by industry provided a basis for weighting the wage series. The report did show, as it turned out, that prices had risen after the controversial McKinley tariff of 1890 had gone into effect, although it also showed that workers had prospered if a longer period was considered. The work for the Aldrich Committee led to further studies of prices by the Bureau of Labor. Beginning in 1902, the Bureau began producing continuing series of wholesale prices and retail food prices. The retail price series used weights from a survey of expenditures by working class families undertaken in 1901 (Weiss 1955, p. 21).

The data produced for the Aldrich Committee was soon put to good use by economists attempting to address the crucial issues of the day. The inflation investigated by Jevons, produced by the discoveries of gold in the 1850s and by the printing of the greenbacks during the Civil War, gave way to deflation as more countries joined the gold standard and massive new supplies of gold could not be found. The deflation, measured by a modern consumer price index, averaged -3.5 percent per year from 1865 to 1879, when the United States returned to the gold standard, and then averaged a milder -1.0 percent per year from 1879 until 1896, when new supplies of gold reversed the trend (Carter et al. 2006, series CC2). The deflation produced a long and bitter political debate. The Populists, eventually led by William Jennings Bryan, claimed that debtors (they were thinking first about farmers) had been hurt. The Populists wanted to switch to a bimetallic monetary regime that would include both gold and silver in the monetary base. The result, they believed, would be an inflation that would reduce the real value of the farm debts that had been made unfairly burdensome by the deflation.

Were the Populist claims valid? The deflation had been ongoing for several decades, and once it was anticipated, it would tend to produce, some economists argued, lower nominal interest rates; farmers who had signed mortgages after deflation was anticipated could not fairly claim that they had been wronged by the gold standard. Moreover, rates on new mortgages would reflect higher expected inflation after a switch to a bimetallic standard. Enter Irving Fisher. In “Appreciation

²Falkner (1899) explains and defends the methodology of the Aldrich Report. Long (1960) discusses the methodology of the report, and compares it with some of the other early federal studies of wages and prices. Conforti (2016) discusses the role of the Aldrich report in the establishment of the wholesale price index.

and Interest” (1896), one of the most influential papers in the history of economics, Fisher explored the relationship between deflation and the rate of interest. Fisher was not the first to propose that nominal interest rates adjust to reflect expected price changes, but he produced the first thorough theoretical and empirical study. Fisher examined a large body of data, relying in part on Jevons’s data for the United Kingdom and the Aldrich Committee’s data for the United States. He found that while the rate of interest had fallen during periods of deflation, the adjustment was slow. As a result, some borrowers had indeed been hurt by the deflation. But Fisher also concluded that most loans then outstanding had been contracted during a time when the downward trend in prices was well-understood, and so for most outstanding loans, there was no injustice to be corrected by inflation.

Later, Fisher (1922) explored the best method of computing price indexes. Fisher reviewed the indexes recommended by Jevons, Laspeyres, Paasche, and others and subjected them to a battery of tests. One concern was what to do about the problem that an index using Laspeyres’s base-year-weighting normally differed from an index using Paasche’s end-year-weighting. Marshall (1887) had suggested the chain-weighted approach that government bureaus have used increasingly in recent years. But when weights are only available at discrete intervals something else must be done. Fisher famously concluded that a geometric average of Laspeyres’s index and Paasche’s was “ideal.”³

An important use of the Bureau of Labor’s price data in a political controversy came during the anthracite coal strike of 1902. The anthracite fields of Pennsylvania provided coal to America’s eastern cities. The strike brought mining to a halt and promised a very cold winter. President Theodore Roosevelt made repeated efforts to mediate. Eventually, he got the two sides to agree to abide by the recommendations of an Anthracite Coal Strike Commission, constructed to balance the interests of the mine owners and the miners. One of the key inputs in the Commission’s work was information about food prices in the mining districts collected by the Bureau of Labor. To provide convincing data, Commissioner Wright flooded the mining districts with agents who collected reams of data. He even authorized the hiring of interpreters because many of the miners were foreign-born. In the end, the Strike Commission’s recommendations appear to have been an attempt to split the difference between labor and capital, but the compilation of data to decide what a “fair” increase in wages would be was both an obvious move by an administration determined to end the strike and an important precedent.

Federal price indexes again became important in World War I. Wages were controlled soon after the war began to hold down the cost of the vast supply of materiel being purchased by the government, but it was understood that wages would need to be raised with the “cost of living.” The retail food price index was pressed into service as a tool for adjusting wages set by government agencies. As early as May 1917, just one month after the United States entered the war, Irving Fisher wrote to

³As a professor I know told her students, “Don’t Laspeyres; if you study hard, you will Paasche the test.”

Assistant Secretary of Labor Louis F. Post suggesting that because food prices by his reckoning rose about twice as fast as prices in general, about one-half of the increase in the food price index would be the appropriate adjustment for wages. Post countered that wages should be raised to the full extent of the increase in the retail food price index (Goldberg and Moye 1985, pp. 102–103).

Clearly, a food price index alone was not a sufficient basis for adjusting wages. The Bureau of Labor Statistics (BLS), as the Bureau of Labor was renamed when it was moved to the newly established Department of Labor in 1913, made its first foray into constructing a more comprehensive index at the request of the Ship Building Labor Adjustment Board, which had the job of setting wages in the shipyards. BLS began by conducting a survey of prices paid by workers in shipbuilding centers. In June 1918, with the National War Labor Board calling for data on the cost of living that covered the entire workforce, and with a substantial emergency grant from President Wilson to finance the effort, the BLS launched a national survey. The resulting report, issued in 1919, was the BLS's first report on the cost of living for the nation as whole. This study provided the weights for the estimates of an index of the "cost of living," which continued to be published after the war (Goldberg and Moye 1985, pp. 102–105).

Interest in the cost of living waned in the prosperous 1920s, and in the 1930s attention focused more on estimates of income and employment. But the return of inflation and wage and price controls in World War II refocused attention on the cost of living index. Labor leaders—George Meany of the AFL was especially vocal—claimed that workers were hurt when the Bureau's index was used to adjust wages because it substantially understated the true increase in the cost of living. Criticism of the index led to the appointment of the President's Committee on the Cost of Living by the War Labor Board. It was a "tripartite committee" with representatives from labor, management, and the public tasked with investigating the index. In January 1944, the labor members issued a stinging indictment of the index. They argued that the true increase in the cost of living between January 1941 and December 1943 was not 23.5 percent as the Cost of Living Index showed, but rather 43.5 percent. What were the errors? The cost of living index, the labor members claimed, failed to take account of the many costs imposed on workers and their families by the war. It did not properly account for the shortage of household workers, the costs of moving to or commuting to jobs in war production centers, higher taxes, quality deterioration, black markets, discontinuance of the production of new consumer durables, rationing, the elimination of sales, and "forced up-trading."

The last item, although unfamiliar today, was an important problem that the Office of Price Administration, the agency that set prices for consumer goods, confronted but never resolved. Manufacturers of products as diverse as underwear and steel normally produced several lines of merchandise: low-quality lines on which the profit margin per unit was low but that the manufacturer could sell in large quantities, and higher-priced lines of higher quality goods sold in smaller quantities but with higher margins per unit. With prices controlled and demand

strong, manufacturers could discontinue their lower-priced line, forcing consumers to “trade up” to the higher-priced line while still claiming that they had not raised their prices. The Office of Price Administration tried various methods for dealing with the problem. For example, it issued regulations that required clothing manufacturers to keep a weighted average of prices of all lines below a maximum, but forced up-trading remained a headache (Rockoff 1984, pp. 151–54).

The claim that the true increase in the cost of living was nearly twice as much as shown by the Bureau of Labor Statistics index led to the appointment of a nonpartisan committee of experts, chaired by Wesley Mitchell of the National Bureau of Economic Research (NBER) and including Simon Kuznets, then at the War Production Board, and Margaret Reid from the Budget Bureau’s Office of Statistical Standards. The committee exonerated the index. It found that the index in December 1943 was understated by about 3 or 4 percent because of unmeasured quality deterioration and by another 0.5 percent because smaller cities were not represented. The committee agreed with labor that the war had imposed many new costs on workers, but doubted that these special war-related costs could be accurately measured, and thought that even if they could be, they should not be included in the cost of living index. The committee did endorse a change in name of the index from “cost of living” to “consumer price,” the name we know it by today (Mitchell 1944; Rockoff 1984, pp. 167–71).

The accuracy of the index receded as a political issue when price and wage controls were removed soon after World War II, but resurfaced when the index began to be used in escalation clauses in contracts. The United Auto Workers, for example, negotiated long-term contracts with wages that escalated with the Consumer Price Index. Labor contended, as it had in the war, that the index understated inflation. For example, the index still did not include taxes. Management, on the other hand, argued that the index failed to take quality improvements fully into account. Although the accuracy of the index was not as pressing an issue as it had been during the war because only part of the labor force was affected; lots of money was at stake. Again, the beleaguered Bureau of Labor Statistics appointed a group of outside experts to investigate the Index and make recommendations. The Price Statistics Commission, popularly known as the Stigler Commission after its chair, economist George Stigler, delivered its report in 1961 (National Bureau of Economic Research, Price Statistics Review Committee 1961).

The Commission was unsympathetic to most of the criticisms that implied the index understated inflation. Taxes are an example. The case for including taxes was straightforward: taxes are the prices paid for services provided by government and should be included just as prices paid for, say, private legal services. Indeed, George Stigler’s famous price theory textbook said exactly that (Stigler 1966, 75 fn. 18). But the Commission for a variety of reasons recommended that taxes be excluded. The Commission was more sympathetic to the notion that the index overstated inflation because it did not take quality improvements fully into account. One of the staff papers written for the Commission was the well-known study by Zvi Griliches (1961) of automobile prices. Griliches, using hedonic methods, estimated the part of the upward

trend in automobile prices that was due to the incorporation of gradual improvements. He concluded that the hedonic method was promising as a research tool, although not yet ready for adoption. The problem of unmeasured quality change due to technological improvement did not end, of course, in the 1960s; it returned with a vengeance with the development of the personal computer and information technology. The Commission's recommendations included establishing a regular schedule of revisions for the weights, introducing new products more promptly, and establishing a research division to develop methods for coping with quality changes.

National Income and Product

There is a long history of attempts to measure national income, dating back at least to the 1665 estimates made by William Petty (Coyle 2014, p. 8).⁴ In the United States, the early estimates made by George Tucker (1843) and Ezra Seaman (1846) have earned high praise from modern students of national income accounting (Gallman 1961). Both relied mainly on census data to produce estimates of population and income, and both sets of estimates were created to address important political issues.

During the depression that followed the Panics of 1837 and 1839, a number of states defaulted on their debts. Tucker (1843, 210–11) argued that this was unnecessary: his state-level estimates of income demonstrated that the states had ample means to repay their debts. Both Tucker and Seaman, moreover, produced revised estimates after the release of the 1850 census. Both also addressed the relative economic strengths of the North and the South, obviously a comparison of crucial importance as the crisis over slavery gathered momentum.

The decisive push for regular publication of estimates of national income came from intense debates over the distribution of income at the turn of the 19th century. Shortly before the United States entered World War I, Scott Nearing (1915) and Wilford Isbel King (1915) published quantitative studies of the distribution of income and wealth. Both studies highlighted the high degree of inequality prevailing in the United States. Nearing's book was not well-received by mainstream economists, who thought it was not done carefully and was influenced too heavily by Nearing's leftist political views. However, King's book was praised—for example by Allyn Young (1916)—and proved to be of enduring influence.

King (1915, pp. 50–1) explained his purpose clearly: "Some of these writers ... contend that the past half century has been an era in which all gains have been absorbed by a few plutocrats while the great masses of the population have become poorer and poorer. Such arguments can only be verified or disproved by a direct study of the facts ..." For data, King relied on the census, as had Tucker and Seaman. But he also utilized death duties from Massachusetts and Wisconsin to construct

⁴Kendrick (1970) provides an international history of attempts to measure national income and product.

Lorenz curves of wealth. These showed that although wealth inequality was great in the United States, it was less than in several industrialized European countries.

Like Nearing (1915), King (1915) estimated the shares of national income going to the factors of production. King found a relatively stable share of rent, refuting the claim that landowners were grabbing a larger and larger share of the national income. The share of rent, moreover, was barely able to cover the expenses of government in 1910, which implied that a single tax on land as proposed by Henry George was infeasible. King (1915, p. 160) did find a sharp decline in the share of wages and salaries from 53.5 percent in 1890 to 46.9 percent in 1910. These numbers might seem to confirm labor's claim that exploitation of labor by capital had increased. But King (p. 163) argued that the decline in labor's share was due to the disappearance of free land in the West and the influx of immigrants of "low efficiency."

King (1915) went on to calculate the real wage per employee. He found a gratifying increase from 1870 to 1900, but then a decrease of about 2.5 percent between 1900 and 1910 (pp. 175–176). He was clear about the cause: not robber barons, but immigrants. King (p. 179) concluded: "The evidence, then, indicates that all the entrenchments of organized labor, all the legislation in favor the working class, all of our new inventions have failed to prevent the invaders from forcing down the commodity wages of American labor." I have dwelled on King, a forgotten figure, not to draw attention to the beliefs and prejudices of an earlier generation of economists, but rather to emphasize that national income accounting grew out of highly charged political controversies over the distribution of income—controversies remarkably similar to those that roil us today.

In 1918, Irving Fisher delivered his presidential address to the American Economic Association. Fisher described the dire condition of the working class, backing up his description with statistics from King and one of his own students. He then argued for tougher inheritance laws and greater participation of the state in the financing and running of industry to reduce inequality. The expansion of the federal government in World War I proved that the government had the capacity to run the economy more efficiently and more equitably than private business. Although Fisher (1919, p. 10) favored a degree of socialism—"we are all socialists now"—he also warned about the danger of supporting extremists who favored class warfare. Fisher's embrace of socialism was not unusual. Republican Warren Harding would easily win the 1920 Presidential election with his call for a return to normalcy, but Socialist Eugene V. Debs would win nearly a million votes—even though he was in jail at the time for urging opposition to the draft.

In the same address, Fisher (1919) suggested an endowment to finance economic research that lay beyond the resources of any one professor. The idea for such an endowment was in the air. For example, a short-lived Bureau of Economic Research had been started in 1899, directed by John R. Commons and financed by George H. Shibley, a wealthy New York lawyer (Bureau of Economic Research 1900). It published a number of quantitative studies, including the first comprehensive index of US stock prices (Bishop 1965). But Shibley was mainly interested in documenting deflation to support his case for bimetallism. Indeed, he wrote a

number of Populist tracts including a 700+ page tome called *The Money Question* (1897). When the price level began to rise at the turn of the 19th century, Shibley lost interest and stopped financing the Bureau.

But Fisher got his wish for a research institution when the National Bureau of Economic Research (NBER) was established. The necessary entrepreneurship had come from a business economist Malcolm C. Rorty and liberal economist Nahum I. Stone, then at the US Tariff Commission. Rorty was impressed by Stone's devotion to objective facts and by Stone's criticism of Nearing's work on the distribution of income (Fabricant 1984). Rorty then assembled a team of economists and business leaders who in June 1917 formed the "Committee on the Distribution of Income," which can be considered the first name of the NBER. World War I delayed further action, but the groundwork had been laid.

The National Bureau of Economic Research was chartered in January 1920. Wesley Clair Mitchell was appointed its first leader, and Mitchell directed its first project: an estimate of national income and its distribution. Mitchell assembled a team that included King, Frederick R. Macaulay, and Oswald W. Knauth. The result was a detailed set of estimates of national income and product that led directly to the modern estimates produced by the federal government. *Income in the United States, Its Amount and Distribution, 1909–1919* was published in two volumes (National Bureau of Economic Research, Mitchell et al. 1921). Nearly 600 pages in all, it far surpassed anything that had come before in terms of the amount of data utilized and the care taken in thoroughly double-checking the component estimates. The study made a clever use of the circular flow. King was tasked with estimating national income from the payments-for-final-products side, and Knauth with estimating it from the payments-for-productive-services side. The two estimates turned out to be reassuringly close. The study identified and tried to deal with many of the problems inherent in estimating national income that continue to be sources of debate and criticism. For example, it noted that its estimate of the national income in 1918 of \$61 billion did not include the monetary value of unpaid work in the household, which probably amounted to "several billions" (National Bureau of Economic Research, Mitchell et al. 1921, p. 143).

Inequality was discussed in the penultimate chapter of the summary volume. This chapter reviewed estimates of the distribution of income by factor of production, estimates of the proportion the population earning less than \$2,000 per year (which is about \$33,000 per year in 2018 dollars inflating with the consumer price index), and estimates by Frederic Macaulay of the personal distribution of income in 1918 based on the newly available income tax returns. The data revealed substantial inequality: the share of income going to the top 1 percent was 14 percent, and the share going to the top 10 percent was 35 percent. The World Inequality Database (at <https://wid.world/>, accessed in September 2018), put these figures for 1918 at similar levels of 16 percent and 40 percent. The chapter on inequality in that first volume, true to the principles of the NBER, does not end with rabble-rousing or policy recommendations. Instead, it ends with a chart and an explanation of a 1918 Lorenz curve.

Although the report itself was just-the-facts, many economists at the time thought that better data could contribute to the design of better legislation. Indeed, Mitchell thought that economic data could do even more. Armed with good data on the state of the economy, including perhaps “leading indicators,” businesses could make wiser decisions about investment and employment. In other words, the business cycle could be tamed with more and better data. The NBER’s estimates of national income and product continued in the 1920s. But they only appeared with a lag and so were of limited value in meeting the fast-changing economic circumstances of an economic depression, as would shortly become clear.

In 1930, Simon Kuznets, a student of Mitchell, took over the Bureau’s project on national income. Kuznets did not invent the concept of national income or national product, nor was he the first to measure them, but he greatly advanced the field. Even a partial list of his achievements must include his creativity in showing how data from a wide array of sources could be combined to build persuasive estimates of national income and product; his investigations of the economic and philosophical issues that bedevil the estimates; his demonstration that many important questions about economic development can be addressed with these estimates (the basis of his Nobel prize); and the role he played as a model scholar-teacher for a generation of economists (Fogel, Fogel, Guglielmo, and Grotte 2013).

The economic catastrophe from 1929 to 1932 produced a US Senate resolution, introduced by Senator Robert La Follette Jr, the Wisconsin Progressive, calling on the Department of Commerce to provide estimates of national income (Dorfman 1959, p. 669). The political point, clearly, was to justify sweeping governmental economic initiatives. Kuznets was seconded to the Department of Commerce to work on an official set of estimates. His first report, submitted in January 1934, showed that national income halved between 1929 and 1932, and although the depth of the Depression was obvious by that time, the report was still an important call to action. President Roosevelt cited the figures, and later cited the updated figures that ran through 1937 when he sent a supplemental budget to Congress in 1938 (Coyle 2014, pp. 12–13).

During World War II, Kuznets joined the War Production Board where he used national income accounting to challenge the spending plans of the War and Navy Departments. Kuznets and Robert Nathan (Kuznets’s student) argued that the spending plans needed to be scaled back, both to prevent competition over supplies from slowing production and to prevent an unacceptable decline in the flow of final goods to consumers. Ultimately, the economists won the “feasibility debate” and forced a cutback in the demands of the Army and Navy. Counterfactual spending plans are hard to evaluate, but two careful students of the feasibility debate have credited the economists with a crucial contribution to the ultimate success of the US munitions program (Edelstein 2011; Lacey 2011).

Kuznets was deeply concerned with the limitations of net national product and related concepts as measures of welfare. He recognized, for example, that improvements in education and healthcare were not measured adequately because they were measured by costs, and he provided some speculative estimates of how the

growth of the net national product was affected by these biases. Ultimately, however, he opted for excluding these speculations from his final estimates.

Military spending presented another problem. In one of his last discussions of national income and product before US entry in World War II, Kuznets (1941, pp. 19–20) explained that his estimates included “dreadnoughts, bombing planes, poison gas, and patent medicines because they are rated economic goods in our country today,” even though they “might well be considered worthless and even harmful” in a society organized differently. In a footnote, Kuznets (p. 31, fn. 5) used an analogy with private spending to buttress his case for including military expenditures: “If the activities of the private police used by many large corporations are productive, why not those of the municipal police? And if of the domestic police, why not of the international police, i.e., the armed forces of the nation?” During World War II, however, Kuznets (1945) modified his thinking. He argued that military spending should be counted in national product during a time of total war, but it should be excluded during peacetime because military spending was then an intermediary good for producing a flow of consumption to consumers. Other economists, including decisively those at the Department of Commerce, thought otherwise (Gilbert, Staehle, Woytinsky, and Kuznets 1944).

A number of economists, however, have found Kuznets’s concept of a Peacetime National Income to be attractive. Higgs (1992), for example, argued that the then-current interpretation of the impact of World War II on the American economy, that it created unprecedented prosperity, was reversed when one used Kuznets’s peacetime concept rather than the conventional measure. Higgs even took exception to Kuznets’s decision to include some military durables such as aircraft in investment because Kuznets thought that they could later be turned to peacetime purposes.

In retrospect, a number of concerns weighed against adopting Kuznets’s concept of peacetime national product. One reason, as Coyle (2014, p. 20) suggests, was the rise of Keynesian economics. In principle, one could use Kuznets’s peacetime version of national product to analyze the macroeconomy, but the conventional measure fit more smoothly into the simple Keynesian model taught to a generation of economics students in Samuelson and other textbooks. Perhaps the most important reason for rejecting Kuznets’s concept, however, was the Cold War. In his famous study of productivity, Kendrick (1961, p. 25) chose to include all defense spending in his estimates of national product partly on the grounds that “national security is *at all times* [Kendrick’s italics] a prime objective of economic organization.” In political terms, excluding national defense from national product would create the appearance that the government’s statistical agency was siding with the critics of America’s defense budget. Of course, no one was required, as Kuznets had pointed out, to use only one measure of aggregate product. To the contrary, Kuznets thought that it would be best to produce a series of measures, some specialized for one purpose and some for another. But as we have learned, public attention does tend to focus on a single measure of national product, so the decision to ignore Kuznets’s peacetime concept may have had important consequences.

Unemployment

The federal unemployment statistics have in common with the price and income statistics that they were born and lived in political controversy. Unlike the price and income statistics, however, the contribution of economists to the thinking about the measurement of unemployment was modest. Their major contribution was mainly in suggesting methods for designing questionnaires and sampling procedures. As Card (2011) put it, the unemployment statistics are more of a case of “Measurement without Theory.”

The first systematic collection of unemployment statistics in the United States seems to have been a survey conducted by the Massachusetts Bureau of the Statistics of Labor in 1878 (Keyssar 1986, pp. 1–5). The head of the Bureau, as noted above, was Carroll D. Wright. Unemployment was running high as a result of the depression that followed the Panic of 1873. Some estimates circulating at the time put the number of unemployed in Massachusetts at 250,000. Wright surveyed local officials who were asked to report on the number of men without work, but to report only those “who really want work,” a definition of unemployment, obviously, that would make for a smaller number than one that included discouraged workers. Wright seems to have been pleased with the results of his survey, which showed at most about 30,000 unemployed workers. This survey appears to be the beginning of an emphasis on people actively seeking work. Close to a century and a half later, the question of whether discouraged workers should be counted as unemployed in the headline estimates of unemployment remains a highly charged political issue.

Questions about unemployment were included in each federal census from 1880 to 1910. The lack of more frequent data was obviously a problem and was felt acutely in 1914 when the outbreak of World War I sparked a financial panic and economic contraction in the United States. The newly renamed Bureau of Labor Statistics (BLS) attempted to estimate the level of unemployment. In 1916, working with agencies in New York, it produced “Unemployment in New York City.” Working with Metropolitan Life Insurance Company (which surveyed its industrial policyholders), it produced “Unemployment in the United States.” The latter study reported a high level of unemployment, which it attributed to several factors including the policy of employers of keeping “40 men in line outside the gates for every job that might open” and the immigrants pouring into the country (Goldberg and Moye 1985, p. 97). In addition, the Bureau of Labor Statistics began tracking and reporting levels of employment in several industries.

Questions about unemployment were dropped from the 1920 census. The timing was unfortunate because the United States suffered a severe “V-shaped” economic contraction in 1920–21. In response, President Harding arranged a conference on unemployment (Hoover 1921). One key question, of course, was the level of unemployment. The conference received contrasting estimates from the Bureau of Labor Statistics and the US Employment Service.

Before looking at the different estimates, a short digression on the Employment Service is warranted, because its history displays the political and economic

forces at work. The idea of publicly funded employment offices to match people seeking work with employers who needed workers had a long history in Europe. An employment service was established in Ohio in 1890, and the idea soon was adopted in other states. The “Ohio idea” moved to the federal level with the establishment of the Division of Information in the Bureau of Immigration and Naturalization in 1907. Net immigration had reached a high of 767,000 in 1907, an addition of nearly 0.9 percent to the population (Carter et al. 2006, series Aa7 and Ad22). Many workers and labor leaders complained that the flow of immigrants was providing strikebreakers and depressing wages. One idea behind the Division of Information was to take the pressure off urban labor markets on the east coast by finding jobs for immigrants in agricultural areas. The Division set up the first federal employment office on Ellis Island in 1907 (Guzda 1983). World War I created an enormous demand for labor. The Division of Information received a large infusion of funds so that it could aid in the placement of war workers; and in recognition of its expanded role, its name was changed to the US Employment Service.

The Hoover Commission received two estimates of unemployment. The Bureau of Labor Statistics estimated the “shrinkage of employment” to be 5.5 million (Goldberg and Moye 1985, p. 126). The BLS emphasized that because it was surveying employers, all it could measure was whether employment had risen or fallen. It didn’t know, for example, how many people were actively seeking work as opposed to how many had decided to leave the labor force. The US Employment Service provided an alternative estimate for September 1921 of 2.3 million for a restricted set of cities based on a survey of state and municipal employment services, employers’ and employees’ organizations, and other sources. The Commission’s Committee on Unemployment Statistics, which included several prominent economists, thought that the Employment Service’s estimate was better, and after extrapolating it to reflect the country as a whole, concluded that the number of unemployed workers was certainly no more and probably less than 3.5 million. The retrospective estimate developed by David Weir (1992), perhaps the best recent estimate, however, is 4.8 million in 1921, much closer to the BLS figure (Carter et al. 2006, series Ba474).

The Committee on Unemployment Statistics made several recommendations for improving the unemployment statistics. For one thing, it was especially skeptical about “reporting over the telephone” (Hoover 1921, p. 40). Interest in the rate of unemployment declined in the prosperous 1920s and funding for the Employment Service dried up.⁵ But another controversy over the level of unemployment erupted soon after the stock market collapse in October 1929. On January 21, 1930, President Hoover announced that information supplied to him by the Department

⁵When Frances Perkins became Secretary of Labor in 1933, the feeling in the administration was that the Employment Service had become merely a letterhead. Consequently, the First US Employment Service was formally abolished in 1933 and a second one created. Later, its programs were absorbed by the Education and Training Administration of the Department of Labor. The employment service still receives funding, but the offices are located with the Department of Labor’s American Job Centers.

of Labor showed that employment was rising (*New York Times*, January 22, 1930). Frances Perkins, New York State's Industrial Commissioner, publicly disagreed: her data from surveys of employers in New York State showed that employment had continued to decline (*New York Times*, January 23, 1930; Perkins 2011, pp. 91–93).

About six weeks later, Hoover reiterated his claim (*New York Times*, March 8, 1930): “The low point of business and employment was the latter part of December and early January. Since that time employment has been slowly increasing, and the situation is much better today than at that time.” To support this claim, Hoover's press release included a memorandum by Robert P. Lamont, the Secretary of Commerce, and James J. Davis, the Secretary of Labor. It included a weekly employment index, which apparently began in mid-December. This index reached a trough of 86.0 on December 30 and then began a steady rise.⁶

Hoover's optimism again received a rebuke from Perkins (*New York Times*, March 11, 1930), who complained that Hoover's weekly figures were not publicly available and guessed that they covered a limited set of firms capable of reporting employment at a weekly frequency, and so were unrepresentative. The publicly available monthly statistics from the Bureau of Labor Statistics and her statistics for New York State showed that the labor market continued to deteriorate. Events soon confirmed Perkins's pessimism. Her willingness to challenge Hoover on this and subsequent occasions impressed Franklin Roosevelt, then governor of New York, which helped to launch Perkins upon the path to becoming the first woman to serve in a President's cabinet.

Questions about unemployment were asked in the 1930 census. The initial findings, published in June 1930, indicated a rate of unemployment of about 6.6 percent. But by the time the data was published, this seemed much too low, and the Census Bureau was asked to do a follow up. This study, conducted in 21 cities, showed unemployment rising from 9.7 percent at the time the 1930 census was conducted to 22.2 percent in January 1931. Other special surveys were conducted during the 1930s, but it had become obvious that something frequent and continuous was needed.

Today we take it for granted that unemployment rates should be estimated monthly by asking a random sample of people questions about whether they are working, and if not, whether they are seeking work, along with additional questions to learn about discouraged workers, part-time work, and related topics. But it was not until 1940 that the Works Projects Administration began a regular monthly survey of employment. The main purpose of the Works Projects Administration was to create jobs, but many of its projects, such as interviewing Americans born into slavery,

⁶ I have not located a source for the weekly statistics. However, an NBER monthly index of automobile production shows increases from the previous month in January, February, March, and April 1930. Also, an NBER monthly index of transportation workers shows an increase from December 1929 to January 1930 (NBER series retrieved from FRED). Thus, it seems possible that Department of Labor might have found some data to support its claims.

turned out to be of enduring value. The unemployment survey is another example. When the Works Projects Administration was terminated in 1942, the unemployment survey was turned over to the Census Bureau. In 1959, responsibility for the survey was transferred to the Bureau of Labor Statistics (Goldberg and Moye 1985, p. 166).

Long after the Depression, economic historians would debate the “true” rate of unemployment during the Great Depression. For some years, the standard estimates were those constructed by Stanley Lebergott (1964). Lebergott’s estimates were challenged in a well-known paper by Michael Darby (1976), who maintained that Lebergott’s decision to count people working for government relief programs produced an overstatement of the true amount of unemployment and understated the rate of recovery in the mid-1930s. In 1936, the year when the two series differ the most, Lebergott’s shows an unemployment rate of 17 percent while Darby’s shows an unemployment rate of 10 percent. Lebergott’s figures were updated by Weir (1992) to take account of Darby’s critique and make other refinements. Margo (1993) provides an overview of this controversy as well as an overview of a number of related issues raised by the unique labor market experience of the 1930s.

The monthly rate of unemployment became a much-watched statistic soon after World War II ended. Many of the criticisms that have dogged the estimates of unemployment in recent years had already surfaced by the late 1950s, such as complaints by the AFL-CIO and others that discouraged workers were not counted as unemployed in the headline number on which the media focused. Taking a longer view, we are still debating who should be counted as unemployed, an issue that influenced the construction of that first unemployment survey conducted by Carroll D. Wright in 1878.

Conclusions

The major federal statistical series are one of the great achievements of economics. True, the federal statistics have not fulfilled the highest hopes that economists had for them such as providing an early warning system for financial panics and economic contractions, as we learned in 2008. Economists have some ideas about where the fault-lines lie: for example, they know that there have been more panics in the United States than in Canada, and they suspect that this will hold in the future (Bordo, Redish, and Rockoff 2015). But economists, like seismologists, cannot yet predict when and where the next earthquake will occur despite the vast amount of data they have collected and studied. Indeed, the regular production of these statistics has not resolved the divisive political debates—such as the debates over the impact of monopolies and immigration on the distribution of income—that called them into being.

A case could be made, I believe, that the federal economic statistics contributed to the development of the macroeconomic policies that were deployed with some success after the last financial panic. We engaged in deficit spending and

monetary expansion based partly on statistics-heavy research by economists. It also seems likely that the federal statistical series have been of considerable value to the private sector. Most private business plans depend in some measure on ideas about where the economy is headed. A manufacturing firm deciding whether to expand production, a financial firm deciding whether to invest in real estate, or a service provider deciding whether to hire more workers, all must make some allowance for the general economic weather. Trends can change unexpectedly, but trends projected from accurate data are more likely to be right than trends based on intuition or the prognostications of uninformed pundits.

Although the role that economic statistics have played in improving the functioning of the economy has been less than was hoped, it seems to me that the development of federal economic statistics has made for less strident and more reasoned debates. Extreme claims about the state of economy can be rejected by pointing to economic statistics produced by the federal government that are widely regarded as sound, if imperfect, and that contribution is clearly important.

■ *I thank Michael Bordo, Eugene White, Julie Day of the Wirtz Labor Library, and the referees and editor for many helpful suggestions. The remaining errors are mine.*

References

- Bishop, George W.** 1965. "A Note on the Origin of the Averaging and Indexing of Stock Prices." *Business History Review* 39(3): 403–05.
- Bordo, Michael D., Angela Redish, and Hugh Rockoff.** 2015. "Why Didn't Canada Have a Banking Crisis in 2008 (or in 1930, or 1907, or ...)?" *Economic History Review* 68(1): 218–43.
- Bureau of Economic Research.** 1900. *Quarterly Bulletin of the Bureau of Economic Research*. Bureau of Economic Research.
- Card, David.** 2011. "Origins of the Unemployment Rate: The Lasting Legacy of Measurement without Theory." *American Economic Review* 101(3): 552–57.
- Carter, Susan, Scott Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright, eds.** 2006. *Historical Statistics of the United States: Millennial Edition*. Cambridge University Press.
- Conforti, Lana.** 2016. "The First 50 Years of the Producer Price Index: Setting Inflation Expectations for Today." *Monthly Labor Review*, June.
- Coyle, Diane.** 2014. *GDP: A Brief but Affectionate History*, Revised and Expanded Edition. Princeton University Press.
- Darby, Michael R.** 1976. "Three-and-a-Half Million U.S. Employees Have Been Misled: Or, an Explanation of Unemployment, 1934–1941." *Journal of Political Economy* 84(1): 1–16.
- Dorfman, Joseph.** 1959. *The Economic Mind in American Civilization: 1606–1933*, 5 vols. New York: Viking Press.
- Duncan, Joseph W., and William C. Shelton.** 1978. *Revolution in United States Government Statistics, 1926–1976*. Washington DC: US Department of Commerce.
- Edelstein, Michael.** 2001. "The Size of the U.S. Armed Forces during World War II: Feasibility and War Planning." In *Research in Economic History*, edited by A. J. Field, G. Clark, and W. A. Sundstrom,

- vol. 20, 47–97. Amsterdam; New York and London.
- Edgeworth, Francis Y.** 1925 [1887, 1889] “Measurement of Change in Value of Money.” In *Papers Relating to Political Economy*, vol. 1., pp. 195–297. London: Macmillan and Co.
- Fabricant, Solomon.** 1984. “Toward a Firmer Basis of Economic Policy.” National Bureau of Economic Research.
- Falkner, Roland P.** 1899. “Wage Statistics in Theory and Practice.” *Publications of the American Statistical Association* 6(46): 275–89.
- Fisher, Irving.** 1896. “Appreciation and Interest.” *Publications of the American Economic Association* 11(4): 1–98.
- Fisher, Irving.** 1919. “Economists in Public Service: Annual Address of the President.” *American Economic Review* 9(1): 5–21.
- Fisher, Irving.** 1922. *The Making of Index Numbers; A Study of their Varieties, Tests, and Reliability*. Houghton Mifflin Company.
- Fogel, Robert William, Enid M. Fogel, Mark Guglielmo, and Nathaniel Grotte.** 2013. *Political Arithmetic: Simon Kuznets and the Empirical Tradition in Economics*. University of Chicago Press.
- Gallman, Robert E.** 1961. “Estimates of American National Product Made before the Civil War.” *Economic Development and Cultural Change* 9(3): 397–412.
- Gilbert, Milton, Hans Staehle, W. S. Woytinsky, and Simon Kuznets.** 1944. “National Product, War and Prewar: Some Comments on Professor Kuznets’ Study and a Reply by Professor Kuznets.” *Review of Economics and Statistics* 26(3): 109–135.
- Goldberg, Joseph P., and William T. Moyo.** 1985. *The First Hundred Years of the Bureau of Labor Statistics*. Bulletin of the United States Bureau of Labor Statistics, 2235. Washington, DC: US Department of Labor.
- Griliches, Zvi.** 1961. “Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change.” In *Price Review Committee, the Price Statistics of the Federal Government*, 173–196. National Bureau of Economic Research.
- Guzda, Henry P.** 1983. “The U.S. Employment Service at 50: It Too Had to Wait Its Turn.” *Monthly Labor Review* 106(6): 12–19.
- Higgs, Robert.** 1992. “Wartime Prosperity? A Reassessment of the U.S. Economy in the 1940s.” *Journal of Economic History* 52(1): 41–60.
- Hoover, Herbert.** 1921. *Report of the President’s Conference on Unemployment*. September 26 to October 13, 1921. Government Printing Office.
- Jevons, W. Stanley.** 1863. *A Serious Fall in the Value of Gold Ascertained, and Its Social Effects Set Forth*. London: E. Stanford.
- Kendrick, John W.** 1961. *Productivity Trends in the United States*. Princeton University Press.
- Kendrick, John W.** 1970. “The Historical Development of National-Income Accounts.” *History of Political Economy* 2(2): 284–315.
- Keyssar, Alexander.** 1986. *Out of Work: The First Century of Unemployment in Massachusetts*. Cambridge University Press.
- King, Willford Isbel.** 1915. *The Wealth and Income of the People of the United States*. Macmillan.
- Kuznets, Simon.** 1941. *National Income and Its Composition, 1919–1938*, 2 vols. National Bureau of Economic Research.
- Kuznets, Simon.** 1945. *National Product in Wartime*. National Bureau of Economic Research.
- Lacey, Jim.** 2011. *Keep from All Thoughtful Men: How U.S. Economists Won World War II*. Annapolis, MD: Naval Institute Press.
- Lebergott, Stanley.** 1964. *Manpower in Economic Growth: The American Record since 1800*. New York: McGraw Hill.
- Long, Clarence D.** 1960. *Wages and Earnings in the United States, 1860–1890*. Princeton University Press.
- Margo, Robert A.** 1993. “Employment and Unemployment in the 1930s.” *Journal of Economic Perspectives* 7(2): 41–59.
- Marshall, Alfred.** 1887. “Remedies for Fluctuations of General Prices.” *Contemporary Review* 51: 355–75.
- Mitchell, Wesley C.** 1938. *The Making and Using of Index Numbers*. Bulletin of the US Bureau of Labor Statistics no. 656.
- Mitchell, Wesley C.** 1944. *Report of the Technical Committee Appointed by the Chairman of the President’s Committee on the Cost of Living*, June 15. Washington.
- National Bureau of Economic Research.** No date. Index of Factory Employment, Transportation for United States [M08F9BUSM331SNBR], retrieved from FRED, Federal Reserve Bank of St. Louis, June 11, 2018.
- National Bureau of Economic Research.** No date. Automobile Production, Passenger Cars, Factory Production for United States [M0107AUSM543NNBR], retrieved from FRED, Federal Reserve Bank of St. Louis, June 11, 2018.
- National Bureau of Economic Research, Price Statistics Review Committee.** 1961. *The Price Statistics of the Federal Government: Review, Appraisal, and Recommendations; A Report to the Office of Statistical Standards, Bureau of The Budget, Together with Twelve Staff Papers*. National Bureau of Economic Research.
- National Bureau of Economic Research (Staff), Wesley C. Mitchell, Willford I. King, Frederick R. Macaulay, and Oswald W. Knauth.** 1921. *Income in the United States: Its Amount and Distribution, 1909–1919*. Edited by Wesley C. Mitchell. National Bureau of Economic Research.

Nearing, Scott. 1915. *Income; an Examination of the Returns for Services Rendered and from Property Owned in the United States*. New York, Macmillan.

New York Times. January 22, 1930. "Employment Turns Upward, Hoover Reports; Changes for First Time since Stock Slump." Special to *The New York Times*, p. 1.

New York Times. January 23, 1930. "Disputes Hover on Employment." p. 11.

New York Times. March 8, 1930. "More Work, Says Hoover, And Depression is Passing; 36 States Are Now Normal." Special to *The New York Times*, p. 1.

New York Times. March 11, 1930. "Employment Drop in State Revealed", p. 2.

Perkins, Frances. [1946] 2011. *The Roosevelt I Knew*. New York: Penguin Books.

Persky, Joseph. 1998. "Retrospectives: Price Indexes and General Exchange Values." *Journal of Economic Perspectives* 12(1): 197–205.

Rockoff, Hugh. 1984. *Drastic Measures: A History of Wage and Price Controls in the United States*. Cambridge University Press.

Seaman, Ezra C. 1846. *Essays on the Progress of Nations, in Productive Industry, Civilization, Population, and Wealth*. New York: Baker and Scribner.

Shibley, George H. 1897. *The Money Question: The 50% Fall in General Prices, the Evil Effects;*

The Remedy, Bimetallism at 16 to 1 and Government Control of Paper Money in Order to Secure a Stable Measure of Prices—Stable Money; Monetary History 1850–1896. Chicago: C.H. Kerr & Company.

Stigler, George J. 1966. *The Theory of Price*, 3rd edition. Macmillan.

Stigler, Stephen M. 1978. "Francis Ysidro Edgeworth, Statistician." *Journal of the Royal Statistical Society. Series A (General)* 141(3): 287–322.

Stigler, Stephen M. 1982. "Jevons as Statistician." *Manchester School* 50(4): 354–65.

Tucker, George. 1843. "Progress of the United States in Population and Wealth in Fifty Years, as Exhibited by the Decennial Census." *Hunt's Merchants Magazine* 9: 2–211.

Weir, David R. 1992. "A Century of U.S. Unemployment, 1890–1990: Revised Estimates and Evidence for Stabilization." In *Research in Economic History*, vol. 14, edited by Roger Ransom, 301–46. Greenwich, CT: JAI Press.

Weiss, Samuel. 1955. "The Development of Index Numbers in the BLS." *Monthly Labor Review* 78(January): 20–5.

World Inequality Database. No date. <https://wid.world/>. Accessed June 1, 2018

Young, Allyn A. 1916. "Nearing's Income; King's Wealth and Income." *Quarterly Journal of Economics* 30(3): 575–87.

Evolving Measurement for an Evolving Economy: Thoughts on 21st Century US Economic Statistics

Ron S. Jarmin

The system of federal economic statistics developed in the 20th century has served the country well, but the current methods for collecting and disseminating these data products are unsustainable. These statistics are heavily reliant on the great 20th century measurement innovation: sample surveys. Recently, however, response rates for both household and business surveys have declined, increasing costs and threatening quality. Existing statistical measures, many developed decades ago, may also miss important aspects of our rapidly evolving economy; moreover, they may not be sufficiently accurate, timely, or granular to meet the increasingly complex needs of data users. Meanwhile, the rapid proliferation of online data and more powerful computation make privacy and confidentiality protections more challenging.

There is broad agreement on the need to transform government statistical agencies from the 20th century survey-centric model to a 21st century model that blends structured survey data with administrative and unstructured alternative digital data sources. For discussions along these lines, see the deliberations of the panel chaired by former Census Bureau Director Robert Groves (National Academies of Sciences 2017), the independent review of the UK statistical system produced by Bean (2016), and the arguments in Bostic, Jarmin, and Moyer (2016). Government statistics in 21st century measurement will be based on vastly more source data, much of which is unstructured—or at least not designed for

■ *Ron S. Jarmin is the Deputy Director and is currently performing the nonexclusive functions and duties of the Director of the US Census Bureau, Washington, DC. His email address is ron.s.jarmin@census.gov.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.33.1.165>

doi=10.1257/jep.33.1.165

statistical uses. Surveys will remain critical, but they will be designed to complement available administrative and alternative data sources and to improve the measurement capabilities of the nontraditional sources. Digital information generated from transactions, online interactions, sensors, the Internet of Things, and many other sources can be used to capture various aspects of economic activity. This change is also likely to result in the decoupling of the collection of basic source data from the processing and dissemination of value-added statistical data products.

The American Economic Association's "Principles of Economic Measurement" (at <https://www.aeaweb.org/content/file?id=6847>) state that economic statistics need to be "Reliable," "Accurate," "Relevant," "Transparent," "Consistent in a Changing World," "Timely," and "Accessible." Traditional survey-based economic statistics are increasingly subject to concerns on all of these dimensions as users place new demands on the data. However, the explosion of digitized information and the tools of modern data science also provide an opportunity for blending traditional survey and new alternative data sources in a way that can lead to the design and delivery of new economic statistics that are more relevant, accurate, timely, and detailed.

In this essay, I describe some work underway that hints at what 21st century official economic measurement will look like and offer some preliminary comments on what is needed to get there. While my focus is on how to improve US economic measurement, there has been substantial progress in other countries that offers lessons for the United States, even in light of differing legal, political, and institutional settings. For example, the Scandinavian countries have made extensive use of linked administrative data, and the Dutch have made impressive and rapid progress moving from survey to administrative and private data sources.

I begin with a brief review of challenges faced by existing government statistical surveys. I continue by highlighting work at the agencies, by academics, and in the private sector that taps new data sources and/or uses new methods to improve economic measurement. I use a variety of examples, but I draw most heavily from activities at the Census Bureau where I have firsthand knowledge. I discuss challenges that will arise in applying new methods of generating statistical data products and sketch a path forward. For example, the use of digital data will depend on the willingness of firms, organizations, and third-party data providers to make available the source data from which more timely and granular economic statistics will be generated. Statistical agencies must maintain production of some current statistical products, but also create new ones in a way that allows longitudinal comparability for at least a select group of measures. Future users of economic statistics are increasingly likely to consume statistical information via specialized products and applications produced by downstream value-added developers in the media, private sector, and academia who will need to work more closely with the agencies to ensure quality.

The Survey Paradigm under Strain

The sample survey has been the workhorse of federal statistical agencies since they were pioneered at the Census Bureau in the middle of the 20th century by Morris Hansen (Hansen, Hurwitz, and Madow 1953). Well-designed and properly executed sample surveys can be an effective method of gathering scientifically robust information. But there are increasing concerns about the viability of surveys. Surveys impose a burden on households and businesses. Moreover, survey organizations in the United States and abroad face secularly declining response rates across a wide range of surveys (described in this journal by Meyer, Mok, and Sullivan 2015; see also Groves 2011; Baruch and Holtom 2008; National Research Council 2013). The downward trend is especially pronounced for the voluntary monthly or quarterly economic indicator surveys, like the Census Bureau's Monthly Retail Trade Survey or the Quarterly Services Survey. To date, the Bureau of Labor Statistics has been able to maintain relatively stable response rates for its business surveys like the Survey of Occupational Injuries and Illnesses or Current Establishment Survey, but like the Census Bureau, its mandatory annual surveys generally outperform voluntary surveys with higher collection frequencies.

These trends will almost certainly continue to be a cause of concern for survey quality and costs. While reliable methods have been developed to deal with missing data in survey estimation (Rubin 1987; Little and Rubin 2002), it is increasingly likely that reliable, unbiased estimation of many critical economic statistics will require new source data.

Limitations to the utility of surveys have been highlighted in a growing number of important contexts. As one example, consider the high-profile issue concerning the number and trends of workers in the so-called "gig economy." Abraham, Haltiwanger, Sandusky, and Spletzer (2017) and Katz and Krueger (2016) compare survey-based data on self-employment from the Current Population Survey, the Survey of Income and Program Participation, and the American Community Survey with administrative data at the aggregate and micro levels. Both sets of authors document that the surveys produce lower estimates of both the size and growth rate of gig economy employment than do the administrative data sources. Farrell and Greig (2017) carry out a big-data exercise on JPMorgan Chase account-holders to document the increase in the number of people receiving income from online platforms. The challenges noted by the Bureau of Labor Statistics in their attempt to measure a type of gig work—"electronically mediated work"—through the Contingent Worker Supplement are further evidence of the difficulty of measuring this part of the economy with survey data alone (Current Population Survey staff 2018).

New Data for Economic Measurement

Surveys are attractive to statistical agencies because they are designed for a specific measurement task. In contrast, administrative and alternative data sources

are designed, collected, and processed for purposes other than economic measurement. Those who wish to repurpose them for statistical uses must seek to understand their measurement properties and their strengths and weaknesses. Struijs, Braaksma, and Daas (2014) discuss the opportunities and challenges of using new alternative data sources for official statistics.

Unsurprisingly, researchers have been faster than the statistical agencies to adapt alternative, and especially government administrative, data to various economic measurement tasks. There has already been a large increase in the utilization of administrative data for research (Chetty 2012) and policy evaluation (Jarmin and O'Hara 2016). Examples include analyses of trends in income equality (for example, Chetty, Hendren, Kline, and Saez 2014) and the changing nature of business dynamics (as in Decker, Haltiwanger, Jarmin, and Miranda 2016). Often, these studies turn to administrative data to study patterns that simply are not available from existing survey-based data—and moreover would be prohibitively expensive to generate in a survey context. In the examples just mentioned, longitudinally linked microdata with universe coverage permit much more precise descriptions of the underlying dynamics than would be possible with survey data.

Importantly, these research efforts can and do lead to innovation in official statistical products. For example, early work on matched employer–employee data (Abowd, Haltiwanger, and Lane 2004) led to the development of the Quarterly Workforce Indicators which integrate many sources of information, including administrative data from state unemployment insurance records and survey-based data from the American Community Survey (Abowd et al. 2009). I discuss additional examples later in the essay.

The use by government statisticians of digital data from private sector sources is also growing, but more slowly. Meanwhile, several companies have developed public-facing data products and tools that can complement official statistics. Examples include the ADP Employment Reports (job growth), Adobe Digital Insights (e-commerce), MasterCard SpendingPulse (consumer spending), and the work of the JPMorgan Chase Institute (bank account transactions). Academic economists have also been occasionally granted access to company data both to address specific research questions and to address broader economic measurement concerns (as in Antenucci, Carafella, Levenstein, Ré, and Shapiro 2014).

One major concern for government statisticians is that they do not have control over the continuity or quality of administrative or other digital data (Landefeld 2014). There is concern that data providers, be they other government agencies (in the case of administrative data) or private sector sources, could disrupt the flow of data to statistical offices or that they may lack the incentive to ensure data quality. Thus, statistical offices around the world have approached the adoption of new data sources cautiously. But large potential payoffs in improved statistical products have generated serious interest in alternative data sources. I focus on some important initial steps at US statistical agencies below.

It is tempting to view these alternative sources of data as substitutes for official statistical products. However, the developers of these alternative sources of data are

quick to note their limitations and their dependence on benchmarking to official data. Thus, it is better to view these private sector and academic products as complements to official statistics. Twenty-first century economic measurement will require that private, academic, and government data providers find new ways to collaborate to leverage their different strengths.

Alternative Data for Measuring the Retail Sector

Recent work at the Census Bureau has focused on augmenting the source data for its monthly and annual statistics on retail trade. The underlying Monthly Retail Trade Survey suffers from low and declining response, dropping from a response rate of 65.9 percent in 2009 to 52.4 percent by 2015. But this work is also motivated by the desire for more timely indicators with more geographical detail than the national estimates available from the survey. Two sources of alternative data are particularly useful in this context: point-of-sale or scanner data and credit/debit card transactions.

Early work on using scanner data to construct prices indices is summarized in Feenstra and Shapiro (2003). More recently, the Census Bureau has been testing the use of data on retail sales obtained from NPD Group to augment its monthly and annual estimates as well as to provide product detail for the 2017 Economic Census. NPD collects point-of-sale data directly from retailers for its market research activities and has negotiated with several large chains to share that information with the Census Bureau. These data track those collected on Census Bureau surveys very closely (Hutchinson 2017). The Census Bureau and NPD are working to expand the number of participating companies with the hope that NPD clients eventually would not need to complete the monthly survey.

One main advantage of data from point-of-sale or credit card transactions is that they may be available at high frequency. The Census Bureau has researched ways to improve its seasonal adjustment procedures using daily credit card transaction data from First Data. These data can help identify and adjust for trading day effects in the Bureau's monthly retail sales indicator releases (McElroy, Monsell, and Hutchinson 2018). These same data made available via a tool built by Palantir have been used by researchers at the Federal Reserve Board to get more precise impacts of disasters like hurricanes on consumer spending (Aladangady et al. 2016).

Neither the point-of-sale nor the credit/debit card data provide complete coverage of the retail sector. That said, their month-to-month changes track official survey data very closely. This might allow a design that employs an annual benchmarking survey coupled with monthly or higher frequency alternative data to construct high-frequency retail sales estimates at subnational levels. However, additional testing with more source data for a larger portion of the retail universe and for more years is required.

In both the NPD and First Data cases, statistical agencies have acquired access to the data for testing purposes through contractual means. It is not clear that a sustainable business arrangement can be negotiated where the agencies could use these as reliable production sources. In the case of NPD, the unit prices that the

Census Bureau is paying for research access would be unaffordable, at least with current budgetary resources, if the project were to be scaled up to cover even the portion of the retail sector captured in NPD data. One possible solution would be if the government allowed third parties like NPD to offer the provision of statistical data to government agencies as part of the service they provide their retail company customers, which would reduce the costs of the survey burden for retail companies. That is, retailers would pay NPD an extra fee to have NPD transmit their retail sales data to the Census Bureau rather than undergoing the burden associated with completing the survey. More broadly, platforms like SAP and QuickBooks could also be useful for facilitating agency access to company data for statistical uses. There are many possible ways to solve the technological, legal, and business relationship issues surrounding the efforts of statistical agencies to get better data in an automated way that minimizes burdens for businesses. Finding the right mix of incentives for all participants—firms, third party data providers, and the statistical agencies—is a challenge to be addressed going forward.

Finally, when the data product in question is a principal economic indicator—like the monthly unemployment rate or retail sales estimate, having the potential to move markets—any partnership with outside providers must take into account the challenge of limiting any reverse-engineering of the indicator. That is, the agencies must be careful that no commercial data provider supplies an identifiable component of the information contained in a principal economic indicator, which would allow that company to predict the statistic more accurately than others before its release.

Alternative Data for Measuring the Healthcare Sector

The Bureau of Economic Analysis has recently introduced a set of Health Care Satellite Accounts (Dunn, Rittmueller, and Whitmire 2015). These satellite accounts depart from the normal procedure of the national income and product accounts by allocating consumer healthcare spending across disease categories, so that the “good” being measured is overall treatment of a disease rather than any of the BEA’s traditional goods and services categories. It allows for the development of disease-specific prices indices, so that data users can separate trends in healthcare spending by the price and quantity of treatment for specific disease categories.

Two approaches are used to construct the Health Care Satellite Accounts. The “MEPS Account” relies on data from the Medical Expenditure Panel Survey (MEPS). The MEPS is the only nationally representative survey on healthcare spending, but its limited sample size can result in noisy estimates. The “Blended Account” augments the MEPS data with billions of claims records (from MarketScan and Medicare) covering millions of individuals. The Blended Account effectively and dramatically increased the sample size for the MEPS, allowing for the smoother and more granular estimates of healthcare spending that researchers and policymakers need. Importantly, it also retains the scientific properties of the MEPS, even though the claims data are not representative of the entire population. This blended survey and alternative data approach will become increasingly important as official statistics integrate more data sources.

Extracting More Value from Existing Data Resources

In some cases, better utilization of the existing data assets can yield significant returns. The diffusion of computing power as early as the 1980s and 1990s allowed researchers to begin to access large micro-datasets in order to examine more fundamental economic units such as establishments, firms, households, individuals, and products. Examples of research in this area include within-industry firm heterogeneity (McGuckin 1993), new goods biases in price indices (Feenstra 1994), and the role of producer dynamics in productivity growth (Foster, Haltiwanger, and Krizan (2001).

The introduction and steady expansion of statistics on business dynamics offers an excellent example of using existing data to generate new statistical products. Following recommendations from the National Academies (Haltiwanger, Lynch, and Mackie 2007), the Census Bureau and the Bureau of Labor Statistics both introduced new statistics based on the longitudinal linkage of their respective business lists. These lists had been used for decades as sampling frames for business surveys and for generating (primarily) cross-sectional statistical products (for example, County Business Patterns). The Census Bureau now publishes the annual Business Dynamics Statistics, and the Bureau of Labor Statistics publishes the quarterly Business Employment Dynamics. While these products are similar, they do have certain key differences that originate in the underlying administrative data that provides the sources for these business lists—the Census Bureau list is based on income and payroll tax data from the IRS, and the Bureau of Labor Statistics list is based on business-level state unemployment insurance data—which in turn can determine their fitness for particular uses (Decker, Haltiwanger, Jarmin, and Miranda 2014). Synchronization of the business lists between the Census Bureau and the Bureau of Labor Statistics would improve data quality, eliminate duplication, reduce the current extraordinary efforts required by Bureau of Economic Analysis to reconcile the statistics derived from the separate business lists, and also enable enhancement and creation of new and improved longitudinal business products (Becker et al. 2005). Differential statutory access to administrative data, particularly to tax data, has hampered these efforts (Shapiro et al. 2014).

These statistical products are entering their second decade, having demonstrated the measurement importance of these universe lists beyond their use as sampling frames. Building on that success, researchers from the Census Bureau and the Federal Reserve recently released a beta product that uses modeling and high-frequency data on applications from firms to the IRS for an “employer identification number,” available with a short lag, to produce more timely data on business startups than is currently available in either the Business Dynamics Statistics or the Business Employment Dynamics. Bayard et al. (2018) describe how to identify “high propensity applications” that are associated with entities that hire employees and grow. The beta version of public-use Business Formation Statistics are available at the national and state level, but the research demonstrates that statistics at the county level and the level of core-based statistical areas (multicounty areas with

an urban center) are feasible as well. Finally, a novel aspect of this work that is not typical for statistical agencies is that the Business Formation Statistics offer projections of future startup activity.

Another recent example of linking two existing data sources is the Opportunity Atlas that combined longitudinal income data from the IRS with Census Bureau data on race and ethnicity. The Opportunity Atlas (<https://www.opportunityatlas.org/>) is a powerful online tool that allows users to see how parental incomes and childhood neighborhood characteristics impact adult outcomes. The methodology for the underlying data is described in Chetty, Friedman, Hendren, Jones, and Porter (2018).

The work represented by the above examples is important not just in providing more timely and granular estimates of business startup activity and income mobility, but also in demonstrating how existing resources can be repurposed to build useful new products. Furthermore, the work highlights the importance of forming teams including subject matter experts who understand the measurement need, those who understand and have access to the data resources, and those with the appropriate technical expertise. The Business Formation Statistics research team is comprised of individuals who possessed decades of experience with Census Bureau surveys and administrative data and are also scholars with a keen sense of measurement needs around business formation. The Opportunity Atlas was possible through the collaboration of academic researchers, Census Bureau experts, and private sector data visualization specialists. The teams these projects rely on for their success possess more diverse skill sets and represent more organizations than is typical for federal survey operations.

Rethinking the Role of Surveys

Given declining survey response rates and the growing availability of alternative data sources, surveys will contribute a declining share of the source data for official statistics over time. Instead of being the primary source data for key economic statistics, survey data should evolve to serve three critical functions.

First, survey data can act as benchmarks to improve the utility of administrative and alternative data. Administrative and alternative data sources often have coverage deficiencies or lack key population characteristics used to make subpopulation estimates like those related to age, race, or industry. For example, the Health Care Satellite Accounts discussed above can use the Medical Expenditure Panel Survey to extend the insights from less systematic data on claims, and the Census Bureau's Small Area Income and Poverty Estimates can use the American Community Survey (ACS) to extend the insights from unrepresentative tax data. It becomes possible to build up estimates on healthcare spending and income, respectively, that are both representative and granular, even though these alternative data on healthcare claims and taxes by themselves lack full coverage and are not representative of the entire US population.

Similarly, preliminary analyses of credit card and point-of-sale data used to generate more timely and granular retail sales estimates indicate that these sources mimic high-frequency changes quite well, but estimates of the level of retail sales can drift away from the official number. Thus, one could envision replacing monthly surveys with a combination of credit card and point-of-sale data that is benchmarked to an annual survey. This would accomplish both a reduction in survey response burden and more timely and granular retail indicators that are benchmarked to scientifically designed survey estimates.

Second, surveys will remain critical for items that are not available from administrative or alternative sources. For example, a relatively new economic collection at the Census Bureau is the Management and Organizational Practices Survey (MOPS) (Buffington, Foster, Jarmin, and Ohlmacher 2017), which, as its name suggests, asks businesses about management practices, something that's not available from company databases. Rather than introducing an entirely new survey, the MOPS is a supplement to the Annual Survey of Manufacturers and thus can use the outcome variables derived from that survey to examine whether differences in management practices translate into business outcomes such as growth, survival, and productivity.

Third, survey design at statistical agencies should over time be optimized to take advantage of the administrative data available to the agencies, as well as new alternative sources. At this point, official statistics on businesses generally make better use of available administrative data than do household surveys. This reflects better availability of administrative data on businesses through the tax and unemployment insurance systems, covering nearly the entire universe of business units. Nevertheless, as highlighted in a recent National Academies of Sciences (2018) report, there remains much scope for better integration of survey and nonsurvey data to improve estimation and to harmonize data and methods across different measurement programs. This optimized survey design involves minimizing the use of surveys to collect information for which administrative or alternative source data are available and prioritizing those items not captured in nonsurvey sources for inclusion in survey data collections.

New Methods throughout the Value Chain of Economic Measurement

Utilizing new sources of data for economic measurement requires statistical agencies to adopt new ways to acquire, ingest, and process these data. The timely and more granular data products we hope to generate from these new sources will also require the agencies to adapt to new ways of marketing and disseminating data to users. Here, I highlight some of the challenges and issues that arise.

Automating Acquisition of Source Data

Statistical agencies have received large amounts of nonsurvey data for decades, but typically via file transfers that require manual mediation. Agencies are looking to automate the large-scale ingestion of nonsurvey data. Passive data collection via

web-scraping is attractive for some data collection tasks. The Bureau of Labor Statistics uses limited web-scraping to supplement its Consumer Price Index program (as discussed in this journal by Groshen, Moyer, Aizcorbe, Bradley, and Friedman 2017) and continues to research ways to passively collect data similar to the Billion Prices Project (discussed in this journal by Cavallo and Rigobon 2016). The Census Bureau is beginning to use web-scraped data to replace survey collections for some state and local government units for its *Quarterly Summary of State and Local Government Tax Revenue* (Dumbacher and Hanna 2017).

Several statistical agencies including the Bureau of Economic Analysis, the Bureau of Labor Statistics, and the Census Bureau are experimenting with acquiring data directly from companies or third parties rather than from surveys. Automated feeds through company “application programming interfaces” (APIs) offer the hope of reducing the burden on survey respondents and improving the quality of source data for the statistical agencies. However, investments need to be made by both businesses and the statistical agencies to make this a reality. Survey responses are designed to meet agency requirements. For example, companies sum sales across products and establishments to report to the monthly retail trade survey. In the future, businesses could transmit more granular summaries of transactions by location and product through third parties or allow access to transactions or summaries of transactions directly via an API. Another possible way statistical agencies could perform statistical computations on company data is through Secure Multi-Party Computing (Goroff 2015). The Census Bureau is currently working with Microsoft on a proof-of-concept to learn how this could be used in the production of economic statistics. Regardless of how the statistical agencies access company data for statistical uses, one critical issue, that will probably need to be resolved on a case-by-case basis, is the degree to which company databases will be required to conform to statistical agency requirements before they are transmitted or otherwise made available to the agencies.

Innovation Requires Putting the Data User First

In the past, the statistical agencies relied on a handful of key (mostly federal government) stakeholders to drive the measurement agenda and to determine what products to develop and publish. The data user community is much broader today and it includes many nongovernment users. Along with traditional economic statistics, many of these users require data that is timelier and more granular. This shift requires government agencies to seek new ways to engage users to identify measurement priorities and to develop products to meet their needs.

The Census Bureau’s Opportunity Project (described at <https://opportunity.census.gov/>) provides a structured and repeatable process to engage stakeholders, the agencies, data scientists, developers, designers, and business strategists around a given measurement problem. Partners have included companies like Cisco and Redfin, nonprofit organizations like the National Urban League, and several government agencies. Initial proofs-of-concept are developed in twelve-week sprints, dramatically faster than typical agency product development timelines. Prototype

products showing promise can be further developed either by agencies or stakeholders; as one example, see the Redfin “Opportunity Score” at <https://labs.redfin.com/opportunity-score>, which shows what types of jobs can be reached from a given location by walking or mass transit, along with information on housing prices in that location.

Increasingly, data users will consume official statistical data via apps that are developed through this type of process. Agencies will need to set up review and quality assurance procedures to facilitate the development of new innovative products. Data users will also consume official economic statistics on platforms over which the agencies have limited control like USAFacts (<https://usafacts.org/>). Ideally, the agencies will work with such platforms to help ensure official products are properly deployed and documented. Clearly, resource constraints will require agencies to prioritize the platforms with which they will choose to work.

Expanding the Toolkit

Continued innovation requires adding the tools of modern data science to the traditional toolkit of the official statistician (National Academies of Sciences 2017). The agencies have responded by updating training to include modern data science methods through partnerships with academia, including the Joint Program on Statistical Methodology. Pilot classes on “Big Data and Federal Statistics” led to the development of a textbook (Foster, Ghani, Jarmin, Krueger, and Lane 2017).

Agency staff are applying modern data science methods within several projects. For example, the Bureau of Labor Statistics and the Census Bureau are exploring how machine learning can be applied to a variety of classification activities (for example, industry and occupation coding). Bertke, Meyers, Wurzelbacker, Measure, Lampl, and Robins (2016) utilize machine learning to improve coding of work-related injuries for the Survey of Occupational Injuries and Illnesses. The Innovation Measurement Initiative, a partnership between the Census Bureau and the University of Michigan’s Institute on Research in Innovation and Science, makes heavy use of the data science toolkit and introduces new data to address a longstanding measurement challenge of relating investments in research and innovation to economic outcomes. Early work is described in Zolas et al. (2015), and the growing data infrastructure is now available to researchers via the Federal Statistical Research Data Centers.

Modernizing Privacy Protections for Granular Data

Providing data users with the more timely and granular estimates while protecting the confidentiality of the underlying data is increasingly difficult for statistical agencies. The proliferation of data on the Internet, along with increased computing power and the availability of software for performing linkages, increase the risk that public statistics may permit re-identification of individuals or businesses in the data from which estimates were tabulated. Importantly, this problem exists for both public use microsamples and aggregated tabulations (Abowd et al. 2017). Academic and private sector researchers have approached this problem from both

statistical and computer science (especially cryptographic) perspectives. Heffetz and Ligett (in this journal, 2014) provide a good introduction to this topic. Fundamentally, no public release from a confidential dataset is completely safe from re-identification. However, differentially private methods provide a way to quantify the privacy risk from publishing a given set of statistics. The essential intuition here is that adding or removing an individual from a dataset should have (nearly) no effect on any publicly released outcome of the data, but accomplishing this requires adding noise to the released data (so that the gain in privacy has a tradeoff of slightly less precision). The Census Bureau's OnTheMap application was the first public data product to employ formal, quantifiable privacy protections (Machanavajjhala, Kifer, Abowd, Gehrke, and Vilhuber 2008).

An excellent example of the need for modern approaches to privacy protection is the new Post-Secondary Employment Outcomes (PSEO) product (described at https://lehd.ces.census.gov/data/pseo_beta.html), which started as a partnership between the Census Bureau and the University of Texas System to develop a tool (available at <https://seekut.utsystem.edu/seekuttool>) for students and their families to examine postgraduation outcomes by campus and degree field. This is accomplished by linking student data from universities and community colleges to workforce data from the Longitudinal Employer-Household Dynamics program at the Census Bureau. This permits the Census Bureau to improve measures of workforce education and within-firm human capital. The privacy protection problem for this application is complicated, because the schools both possess the original lists and in some cases publish statistics from them. The solution infuses noise into the data, in a way that provides "provable differential privacy" (US Census Bureau 2018) and allows users to generate queries that would otherwise have high risk of disclosing data for particular students. Given the initial success of the pilot PSEO, the Census Bureau is in the process of adding schools to scale up the project.

Path Forward

It is possible to envision a future in which there is more private sector and/or academic provision of economic statistics. ADP, the JPMorgan Chase Institute, and the Billion Prices Project demonstrate that high-quality statistics based on alternative data sources can be produced outside of the statistical agencies. While these developments are welcome, there are two reasons why federal statistical agencies will still need to play a central role in the provision of economic statistics.

First, although the information underlying them may not be so, official statistics are essentially public goods. Socially efficient provision of statistical information on the structure, evolution, and performance of the economy requires the resources and the coordination function of the federal government. The federal government also has the ability to provide continuity, and it is difficult to envision the scale and scope of statistical production being fully duplicated by the private or academic sectors.

Second, private sector provision may not always enjoy the same degree of public trust as official statistics, because private actors may have incentives to use pre-release information for private gain. Thus, the official statistical agencies need to work with private sector and academic providers to identify opportunities where they can add value along some quality dimension that the agencies find difficult, which may be especially relevant for big data that is more timely or detailed but perhaps not representative of the population of interest. In this way, efficient provisioning of publicly available official statistics requires planning and coordination as well as a commitment by the statistical agencies to sound methods and transparency.

Access and Data Governance

Some foreign statistical offices have made substantial progress using alternate data sources and methods. For example, Statistics Netherlands created estimates of traffic intensities in Holland using GPS data and traffic sensors (Ma, van Dalen, de Blois, and Kroon 2011) and Belgian daytime population estimates have been constructed using mobile phone data (Deville et al. 2014). Of course, legal and institutional characteristics may constrain access and use differently across countries, but the US statistical agencies should pay attention to innovations in other countries.

A key question is whether some statutorily mandated US data collections, like the Economic Census, could be done through automated data access modes (like application programming interfaces), or via secure multiparty computing. If so, and if the additional reporting burden for companies was minimal, could the frequency and quantity of information delivered be increased? What sort of legal and policy structure is needed to govern how statistical agencies access private data assets for statistical uses? Another key question is whether companies would seek to charge for access to data delivered in this way. If so, would the associated fees be within the budgetary resources of the statistical agencies or other public-spirited organizations committed to high-quality economic statistics? Similarly, federal, state, and local government agencies that possess important administrative data are not always incentivized to make them available for statistical uses. The recommendations of the Commission on Evidence-Based Policymaking (2017) and the Federal Data Strategy (<https://strategy.data.gov/>) seek to encourage agencies to permit secure and lawful access for statistical, evaluative, and research purposes. Perhaps a bigger concern is how changes in, for instance, tax law policy could affect the content or coverage of administrative data available for statistical purposes.

Methods and Processes

As blended data from surveys and alternative sources become the norm for government statistics rather than the exception, the way in which such estimates are modeled will play a much larger role. The statistical agencies have some experience with blended data from programs such as the Census Bureau's Small Area Income and Poverty Estimates using methods based on Fay and Herriot (1979). Lohr and Raghunathan (2017) provide an updated review of methods for combining survey and nonsurvey data.

In addition to producing estimates from blended data, the statistical agencies must expand their general data science capabilities. Varian (in this journal, 2014) discusses modern data science tools, many of which can be applied to generating economic statistics from the types of data discussed here. An important point of focus for the statistical agencies will be data acquisition and curation as new unstructured data sources are accessed, tested, and applied to measurement problems. The ingestion, processing, and curation of large administrative and alternative sources introduces scalability concerns not present in most survey contexts. Negotiating access will require the agencies to get more staff with skills to initiate and manage business relationships with data providers.

Modernized confidentiality protection also has methodological and computational implications. The statistical agencies currently have the computer science expertise to scale “formally private” disclosure protection methods for only a small number of products. The Census Bureau is the only agency to implement these procedures in production, and its team is currently fully tasked in modernizing its confidentiality systems for the 2020 Census. Relatedly, cryptographic methods like secure multiparty computing might allow statistical agencies to perform privacy-preserving computations, but the agencies currently do not have enough expertise to move beyond small-scale testing. As yet, statistical agencies lack the technical expertise to fully test and scale solutions for these promising technologies.

Clearly, modernization requires significant investments in computer science and engineering expertise at the statistical agencies. This is a major challenge given the competition for attracting talent across other government agencies and the private sector. Collaboration with academic experts and contracting can be part of the solution, but some internal expertise is essential.

In general, agencies should seek to expand substantially the accessibility and usability of their data products. In the future, many users will access statistical information through special purpose apps (whether provisioned by the agencies themselves or by others relying on the application programming interfaces from the agencies), which requires more user-centric approaches to the design and provisioning of agency data products. This will require difficult resource allocation decisions as most statistical agencies concentrate the bulk of their resources on collection and processing, not dissemination.

Priorities, Duplication, and Organization

The current federal statistical system is decentralized. There are 13 principal statistical agencies, with three that are most heavily involved in economic measurement: the Bureau of Economic Analysis, Bureau of Labor Statistics, and the Census Bureau. A decentralized system does not impose high additional costs when most economic statistics are produced within vertically integrated stand-alone surveys, or where data linkage is conducted at an aggregated level. But a system that takes advantage of linking surveys, administrative, and third-party data at the micro-level, and where data are captured via application programming interfaces, scraped from the web or harvested

from sensors, could gain from leveraging scale economies in data access and processing.

Many commenters on this issue, including former Commissioner of Labor Statistics Janet Norwood (1995), have contemplated reorganizing the federal statistical system. The Confidential Information Protection and Statistical Efficiency Act of 2002 called for greater collaboration and business data sharing between the Bureau of Economic Analysis, the Bureau of Labor Statistics, and the Census Bureau, but the Census Bureau is limited in sharing data that are commingled with federal tax information (Pilot 2011). Recently, the Office of Management and Budget (2018) proposed moving the Bureau of Labor Statistics to the Commerce Department, which is already home to the Bureau of Economic Analysis and the Census Bureau. This is a positive step, but resolving all legal issues around data sharing would probably require a full merger of the three agencies. Congressional approval would be required, whether the three agencies are merged or whether BLS is moved to the Commerce Department.

Regardless of whether the federal statistical agencies are realigned or reorganized, they need to review their current structure given how the production of statistics is evolving. Current structures tend to be vertically integrated and centered on specific data collection activities. This structure results in some activities being starved for talent, while others do not fully utilize skilled staff. The Census Bureau has made partial progress on organizing more functionally with matrixed multidisciplinary teams, but more is needed to allocate staff to agency priorities most effectively.

The collective economic measurement system will need to make a number of investments. It will need to invest in building relationships across government agencies and the private sector to secure access to high-quality source data. It will need to invest in staff with the skills to acquire, process, and curate large datasets, and to build reliable and privacy-protected statistical products from blended data. Information systems need to be redesigned to accommodate both survey and alternative data processing.

The Role of Economists and Social Scientists

The statistical agencies will need to work with many external partners to provide high-quality 21st-century economic statistics. The need for combined expertise across subject matter areas, including economics, demography, statistics, data science, computer science, and others, suggests that interdisciplinary teams will be important to improving economic measurement. Indeed, many of the new products developed with administrative and alternative data described above were developed by economists and other social scientists pursuing measurement-focused, curiosity-driven research—some directly employed by the agencies, some from outside government in collaboration with agency staff. Academic economists have long worked with the federal statistical system to improve economic measurement, especially through the Conference on Research in Income and Wealth (<http://www.nber.org/CRIW/>). I encourage academic economists with interest in measurement to get more directly involved with the agencies and the CRIW.

Recent years have seen a large increase in economists working in companies, especially tech companies. Given their joint interest in using new data sources for various aspects of economic measurement, there is a natural affinity between economists in the statistical agencies and those working in the private sector. The National Association for Business Economics (NABE) has been providing important opportunities for these economists to interact and explore ways to work together. An important note regarding this for academic departments training economists is that a growing share of PhD economists are now working in multidisciplinary settings where the goal is not always publishing in the scholarly literature. How can we give those not taking academic jobs more exposure to the methods and language of diverse disciplines like computer science, product design, and so on?

Agencies must continue to improve access to confidential data for qualified researchers such as through secure systems like the Federal Statistical Research Data Centers. The issues surrounding research and program evaluation use of confidential statistical and administrative data are discussed in Foster, Jarmin, and Riggs (2009) and the Commission on Evidence-Based Policymaking (2017). Secure access systems like the Federal Statistical Research Data Centers enhance agency collaboration with academic researchers, facilitate knowledge transfer, and leverage the value of nonpublic data assets.

Conclusion

The system of economic measurement developed in the 20th century continues to provide critical statistics on the health and performance of the economy. That said, current measurement programs are not keeping pace with the changing economy, and current methods for collecting and disseminating statistical information are not sustainable. A process is underway in which government statisticians will take greater advantage of the proliferation of new data sources that can be utilized for economic measurement, as well as new tools for processing, analyzing, and publishing improved economic statistics. My hope is this brief summary can begin a conversation both within the economics profession and across the entire set of stakeholders on how to build a 21st-century economic measurement system.

■ *I would like to thank John Abowd, Katharine Abraham, Catherine Buffington, Lucia Foster, John Haltiwanger, Timothy Taylor, and the editors of the JEP for helpful comments. Some of work described here was supported by the Kauffman and Sloan Foundations. The views expressed in this article are those of the author and not necessarily those of the US Census Bureau.*

References

- Abowd, John, Lorenzo Alvisi, Cynthia Dwork, Sampath Kannan, Ashwin Machanavajhala, and Jerome Reiter. 2017. "Privacy-Preserving Data Analysis for the Federal Statistical Agencies." January. arXiv:1701.00752
- Abowd, John M., John Haltiwanger, and Julia Lane. 2004. "Integrated Longitudinal Employer-Employee Data for the United States." *American Economic Review* 94(2): 224–29.
- Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." In *Producer Dynamics: New Evidence from Micro Data*, edited by Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, 149–230. University of Chicago Press.
- Abraham, Katharine, John Haltiwanger, Kristin Sandusky, and James Spletzer. 2017. "Measuring the Gig Economy: Current Knowledge and Open Issues." (Forthcoming in *CRIW: Measuring and Accounting for Innovation in the 21st Century*. University of Chicago Press.) <http://www.nber.org/chapters/c13887>.
- Aladangady, Aditya, Shifrah Aron-Dine, Wendy Dunn, Laura Feiveson, Paul Lengermann, and Claudia Sahn. 2016. "The Effect of Hurricane Matthew on Consumer Spending." *FEDS Notes*, December 2, Board of Governors of the Federal Reserve System (US).
- Antenucci, Dolan, Michael Cafarella, Margaret Levenstein, Christopher Ré, and Matthew D. Shapiro. 2014. "Using Social Media to Measure Labor Market Flows," March. NBER Working Paper 20010.
- Baruch, Yehuda, and Brooks C. Holtom. 2008. "Survey Response Rate Levels and Trends in Organizational Research." *Human Relations* 61(8): 1139–60.
- Bayard, Kimberly, Emin Dinlersoz, Timothy Dunne, John Haltiwanger, Javier Miranda, and John Stevens. 2018. "Early-Stage Business Formation: An Analysis of Applications for Employer Identification Numbers." NBER Working Paper 24364.
- Bean, Charles. 2016. Independent Review of UK Economic Statistics. March. GOV.UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf.
- Becker, Randy, Joel Elvery, Lucia Foster, C. J. Krizan, Sang Nguyen, and David Talan. 2005. "A Comparison of the Business Registers Used by the Bureau of Labor Statistics and the Bureau of the Census." August. <https://www.bls.gov/osmr/pdf/st050270.pdf>.
- Bertke, S. J., A. R. Meyers, S. J. Wurzelbacher, A. Measure, M. P. Lampl, and D. Robins. 2016. "Comparison of Methods for Auto-Coding Causation of Injury Narratives." *Accident Analysis & Prevention* 88: 117–23.
- Bostic, William G., Jr., Ron S. Jarmin, and Brian Moyer. 2016. "Modernizing Federal Economic Statistics." *American Economic Review* 106(5): 161–64.
- Buffington, Catherine, Lucia Foster, Ron Jarmin, and Scott Ohlmacher. 2017. "The Management and Organizational Practices Survey (MOPS): An Overview." *Journal of Economic and Social Measurement* 42(1): 1–26.
- Current Population Survey staff. 2018. "Electronically Mediated Work: New Questions in the Contingent Worker Supplement." *Monthly Labor Review*, US Bureau of Labor Statistics, September.
- Cavallo, Alberto, and Roberto Rigobon. 2016. "The Billion Prices Project: Using Online Prices for Measurement and Research." *Journal of Economic Perspectives* 30(2): 151–78.
- Chetty, Raj. 2012. "Time Trends in the Use of Administrative Data for Empirical Research." https://www.rajchetty.com/chettyfiles/admin_data_trends.pdf.
- Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2018. "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility." Working Papers 18-42, Center for Economic Studies, US Census Bureau.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics* 129(4): 1553–1623.
- Commission on Evidence-Based Policymaking. 2017. *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking*. September. <https://www.cep.gov/content/dam/cep/report/cep-final-report.pdf>.
- Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The Role of Entrepreneurship in US Job Creation and Economic Dynamism." *Journal of Economic Perspectives* 28(3): 3–24.
- Decker, Ryan A., John C. Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2016. "Where Has All

the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S.” *European Economic Review* 86: 4–23.

Deville, Pierre, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R. Stevens, Andrea E. Gaughan, Vincent D. Blondel, and Andrew J. Tatem. 2014. “Dynamic Population Mapping Using Mobile Phone Data.” *PNAS* 111(45): 15888–93.

Dumbacher, Brian Arthur, and Demetria Hanna. 2017. “Using Passive Data Collection, System-to-System Data Collection, and Machine Learning to Improve Economic Surveys.” JSM 2017 Online Program, Baltimore, MD. <http://ww2.amstat.org/meetings/jsm/2017/onlineprogram/AbstractDetails.cfm?abstractid=322018>.

Dunn, Abe, Lindsey Rittmueller, and Bryn Whitmire. 2015. “Introducing the New BEA Health Care Satellite Account.” *Survey of Current Business*, January. https://www.bea.gov/scb/pdf/2015/01January/0115_bea_health_care_satellite_account.pdf.

Farrell, Diana, and Fiona E. Greig. 2017. “The Online Platform Economy: Has Growth Peaked?” Available at SSRN: <https://ssrn.com/abstract=2911194>.

Fay, Robert E., III, and Roger A. Herriot. 1979. “Estimates of Income for Small Places: An Application of James–Stein Procedures to Census Data.” *Journal of the American Statistical Association* 74(366): 269–77.

Feenstra, Robert C. 1994. “New Product Varieties and the Measurement of International Prices.” *American Economic Review* 84(1): 157–77.

Feenstra, Robert C., and Matthew D. Shapiro, eds. 2003. *Scanner Data and Price Indexes*. University of Chicago Press.

Foster, Ian, Rayid Ghani, Ron S. Jarmin, Frauke Kreuter, and Julia Lane, eds. 2017. *Big Data and Social Science: A Practical Guide to Methods and Tools*. CRC Press.

Foster, Lucia, John Haltiwanger, and C. J. Krizan. 2001. “Aggregate Productivity Growth: Lessons from Microeconomic Evidence.” Chap. 8. in *New Developments in Productivity Analysis*, edited by Charles R. Hulten, Edwin R. Dean, and Michael J. Harper. University of Chicago Press.

Foster, Lucia, Ron Jarmin, and Lynn Riggs. 2009. “Resolving the Tension between Access and Confidentiality: Past Experience and Future Plans at the U.S. Census Bureau.” *Statistical Journal of the IAOS* 26(3–4): 113–22.

Goroff, Daniel L. 2015. “Information Access. Balancing Privacy versus Accuracy in Research Protocols.” *Science*, January 23, 347(6221): 479–80.

Groshen, Erica L., Brian C. Moyer, Ana M.

Aizcorbe, Ralph Bradley, and David M. Friedman. 2017. “How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches.” *Journal of Economic Perspectives* 31(2): 187–210.

Groves, Robert M. 2011. “Three Eras of Survey Research.” *Public Opinion Quarterly* 75(5): 861–71.

Haltiwanger, John, Lisa M. Lynch, and Christopher Mackie, eds. 2007. *Understanding Business Dynamics: An Integrated Data System for America’s Future*. National Academy Press.

Hansen, Morris H., William N. Hurwitz, and William G. Madow. 1953. *Sample Survey Methods and Theory*, vols. 1 and 2. New York: John Wiley & Sons.

Heffetz, Ori, and Katrina Ligett. 2014. “Privacy and Data-Based Research.” *Journal of Economic Perspectives* 28(2): 75–98.

Hutchinson, Rebecca J. 2017. “Reducing Survey Burden Through Third-Party Data Sources.” Conference of European Statisticians, Workshop on Statistical Data Collection, October 10–12, 2017, Ottawa, Canada, WP. 1-4. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2017/mtg3/DC2017_1-4_Hutchinson_USA_AD.pdf.

Jarmin, Ron S., and Amy B. O’Hara. 2016. “Big Data and the Transformation of Public Policy Analysis.” *Journal of Policy Analysis and Management* 35(3): 715–21.

Katz, Lawrence F., and Alan B. Krueger. 2016. “The Rise and Nature of Alternative Work Arrangements in the United States, 1995–2015.” NBER Working Paper 22667.

Landefeld, Steve. 2014. “Uses of Big Data for Official Statistics: Privacy, Incentives, Statistical Challenges, and Other Issues.” <https://unstats.un.org/unsd/trade/events/2014/beijing/Steve%20Landefeld%20-%20Uses%20of%20Big%20Data%20for%20official%20statistics.pdf>.

Little, Roderick J. A., and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*. Hoboken NJ: John Wiley & Sons.

Lohr, Sharon L., and Trivellore E. Raghunathan. 2017. “Combining Survey Data with Other Data Sources.” *Statistical Science* 32(2): 293–312.

Ma, Yinyi, Jan van Dalen, Chris de Blois, and Leo Kroon. 2011. “Estimation of Dynamic Traffic Densities for Official Statistics.” *Transportation Research Record: Journal of the Transportation Research Board* 2256(1): 104–11.

Machanavajjhala, Ashwin, Daniel Kifer, John Abowd, Johannes Gehrke, and Lars Vilhuber. 2008. “Privacy: Theory Meets Practice on the Map.” In *IDCE ‘08: Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, 277–86.

ACM Digital Library.

McElroy, Tucker S., Brian C. Monsell, and Rebecca Hutchinson. 2018. "Modeling of Holiday Effects and Seasonality in Daily Time Series." Center for Statistical Research and Methodology Report Series RRS2018/01. Washington: US Census Bureau.

McGuckin, Robert H. 1993. "The Importance of Establishment Data in Economic Research." Working Papers 93-10, Center for Economic Studies, U.S. Census Bureau.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29(4): 199–226.

National Academies of Sciences, Engineering, and Medicine. 2017. *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*, edited by Robert M. Groves and Brian A. Harris-Kojetin. Washington, DC: National Academies Press.

National Academies of Sciences, Engineering, and Medicine. 2018. *Reengineering the Census Bureau's Annual Economic Surveys*. Edited by Katharine Abraham, Constance F. Citro, Glenn D. White, and Nancy K. Kirkendall. Washington, DC: National Academies Press.

National Research Council. 2013. *Nonresponse in Social Science Surveys: A Research Agenda*. Washington, DC: National Academies Press.

Norwood, Janet L. 1995. *Organizing to Count: Change in the Federal Statistical System*. Urban Institute Press.

Office of Management and Budget, Executive Office of the President of the United States. 2018. "Delivering Government Solutions in the 21st Century: Reform Plan and Reorganization

Recommendations." <https://www.whitehouse.gov/wp-content/uploads/2018/06/Government-Reform-and-Reorg-Plan.pdf>.

Pilot, Adrienne. 2011. "Data Synchronization: Leveraging Existing Business Data to Better Measure the Economy." *Amstat News*, November 1. <http://magazine.amstat.org/blog/2011/11/01/data-synchronizationscipolicy/>.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Shapiro, Matthew, Barry P. Bosworth, Don A. Dillman, John C. Haltiwanger, Colm O'Muircheartaigh, Richard D. Rippe, and Kirk M. Wolter. 2014. "A Statement on Data Synchronization from the Members of the Federal Economic Statistics Advisory Committee (FESAC)." FESAC Statement on Data Synchronization." https://www2.census.gov/adrm/fesac/2014-12-12_fesac_data_sync.pdf.

Struijs, Peter, Barteld Braaksma, and Piet J. H. Daas. 2014. "Official Statistics and Big Data." *Big Data & Society* 1(1).

US Census Bureau. 2018. "Technical Appendix for Protection System Post-Secondary Employment Outcomes (PSEO) (Beta)." https://lehd.ces.census.gov/doc/pseo_appendix_tech_doc.pdf.

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28(2): 3–28.

Zolas, Nikolas, Nathan Goldschlag, Ron Jarmin, Paula Stephan, Jason Owen-Smith, Rebecca F. Rosen, Barbara McFadden Allen, Bruce A. Weinberg, and Julia I. Lane. 2015. "Wrapping It Up in a Person: Examining Employment and Earnings Outcomes for Ph.D. Recipients." *Science*, December 11, 350(6266): 1367–71.

Environmental Justice: The Economics of Race, Place, and Pollution

Spencer Banzhaf, Lala Ma, and Christopher Timmins

In 1978, 31,000 gallons of polychlorinated biphenyl (PCB)—a highly dangerous chemical—were illegally dumped on behalf of the Ward Transformer Company across 14 counties in North Carolina. The state collected the contaminated soil and identified two landfill sites for the waste: a publicly owned landfill in Chatham County, and a recently foreclosed property in Warren County. The Warren County site had a shallow water table, making it unsuitable for a landfill. However, the site was privately owned and near a town with no mayor or city council. In contrast, the Chatham County site was publicly owned, giving local residents an opportunity to participate in the siting decision. Additionally, in 1980, Warren County was 60 percent black and 25 percent of its families were below the poverty line (and the area immediately near the site had even higher proportions of people of color), whereas the corresponding figures for Chatham County were only 27 percent and 6 percent. Ultimately, the state placed the landfill in Warren County. Protests over this decision drew widespread support from civil rights groups and gained national media attention. The environmental justice movement was born (for a detailed description of this episode, see Hampson 2010).

■ *Spencer Banzhaf is Professor of Economics, Andrew Young School of Policy Studies, Georgia State University, Atlanta, Georgia. Lala Ma is Assistant Professor of Economics, Gatton College of Business and Economics, University of Kentucky, Lexington, Kentucky. Christopher Timmins is Professor of Economics, Duke University, Durham, North Carolina. Their email addresses are hsbanzhaf@gsu.edu, lala.ma@uky.edu, and christopher.timmins@duke.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.33.1.185>

doi=10.1257/jep.33.1.185

The grassroots movement that placed environmental justice issues on the national stage was soon followed up by research documenting the correlation between pollution and race and poverty (Bullard 1983; US GAO 1983; Commission for Racial Justice, United Church of Christ 1987). Since then, a growing body of academic work in law, sociology, public policy, geosciences, and economics has further investigated such correlations under a variety of measures of exposure, spatial scales, and statistical controls. This work has established inequitable exposure to nuisances as a stylized fact of social science. Useful overviews of this literature include Bullard (1994), Cole and Foster (2001), Bowen (2002), Ringquist (2003, 2005), Noonan (2008), Mohai, Pellow, and Roberts (2009), and Banzhaf (2012).

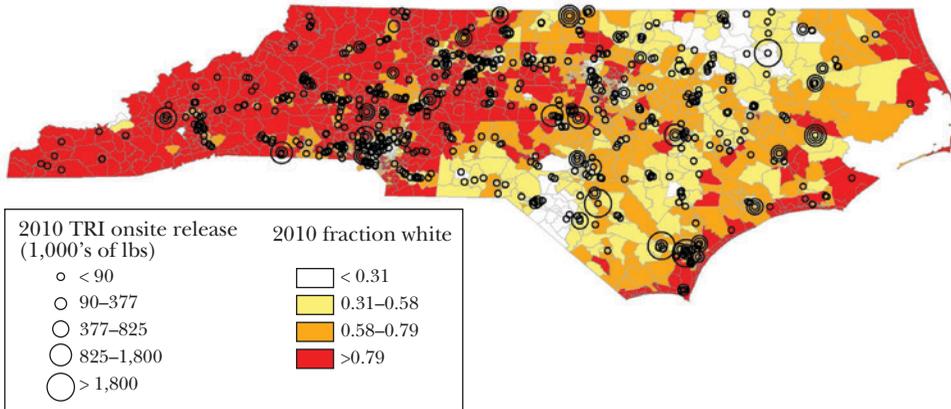
Figures 1–3 offer an example of the evidence behind such correlations, in this case the correlation between demographics and the location of large polluters identified in the Toxic Release Inventory (TRI), a national database established by law which requires private and government facilities to report annually how much of certain chemicals they emit into air or water or send to landfills. Figure 1 plots emissions from these polluters in North Carolina in the year 2010 with circles, against a heat map of the percentage of the population that is white at the census tract level. The correlation between the population density of people of color and emissions intensities at these facilities is striking. A simple calculation finds that the share of 2010 sites operating in tracts that are more than 80 percent non-Hispanic white is about double the share in tracts that are more than 80 percent non-white (22 versus 10 percent). Yet, there are almost five times as many mostly white tracts as there are tracts predominantly of people of color (788 versus 166). Figure 2 next maps emissions against income at the county level. Zooming in on a particular area of Figure 2, Figure 3 shows that facilities are disproportionately located in the lower-income tracts within the county. In this case, the share of TRI facilities operating in tracts with per capita income below \$21,000 is 63 percent.

Such correlations indicate, but do not fully encapsulate, the concerns of the environmental justice movement. These concerns pertain both to distributive justice (relating to the distribution of environmental burdens) and to procedural justice (relating to the decision-making processes that lead to those distributions). For example, the US Environmental Protection Agency defines “environmental justice” as requiring that “no group of people should bear a disproportionate share of the negative environmental consequences resulting from industrial, governmental and commercial operations or policies” and calls for “fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation and enforcement of environmental laws, regulations and policies.”¹ More generally, environmental justice dovetails with the growing concern about income inequality (for example, Chetty, Hendren, Kline, Saez, and Turner 2014; Piketty 2014). Because public goods are part of

¹See the EPA webpage “Learning about Environmental Justice” at <https://www.epa.gov/environmentaljustice/learn-about-environmental-justice>.

Figure 1

Emissions from Large Polluters and Fraction Non-Hispanic White for North Carolina, 2010

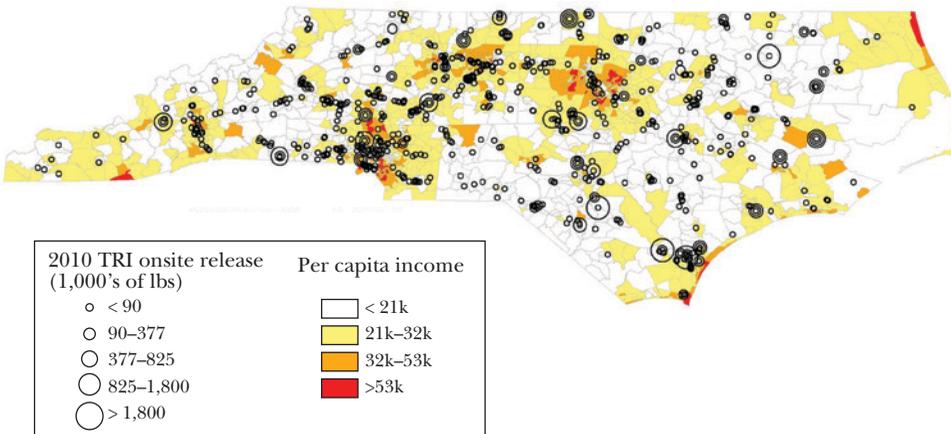


Source: Authors using data from the Toxic Release Inventory and US Census.

Note: Using data from the Toxic Release Inventory (TRI), Figure 1 plots emissions from large polluters in North Carolina in the year 2010 with circles, against a heat map of the percentage of the population that is non-Hispanic white at the census tract level.

Figure 2

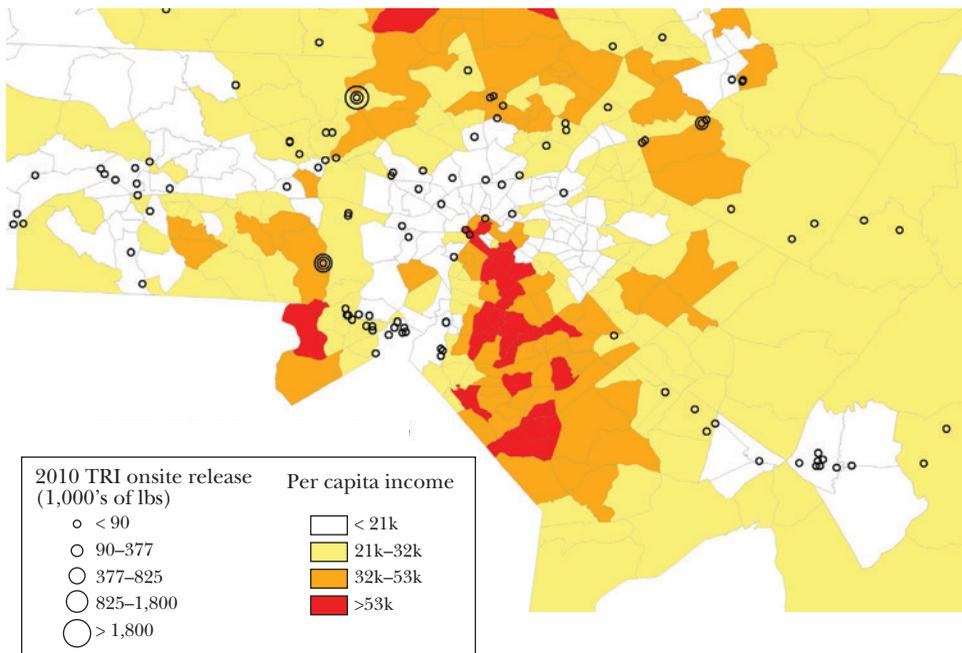
Emissions from Large Polluters and Per Capita Income for North Carolina, 2010



Source: Authors using data from the Toxic Release Inventory and US Census.

Note: Using data from the Toxic Release Inventory (TRI), Figure 2 plots emissions from large polluters in North Carolina in the year 2010 with circles, against a heat map of income at the census tract level.

Figure 3

Emissions from Large Polluters and Per Capita Income, Part of North Carolina, 2010

Source: Authors using data from the Toxic Release Inventory and US Census.

Note: Using data from the Toxic Release Inventory (TRI), Figure 3 plots emissions from large polluters in a part of North Carolina in the year 2010 with circles, against a heat map of income at the census tract level.

households' real income, the distribution of environmental amenities is part of the overall landscape of inequality.

In this paper, we review the environmental justice literature, especially where it intersects with work by economists. Although environmental justice is an interdisciplinary field, economists' focus on causal relationships, and on linking empirical models to theoretical ones, gives them a comparative advantage in untangling the web of socioeconomic relationships involved. In the next section, we first consider in more depth the literature documenting evidence of disproportionate exposure, such as that offered by Figures 1–3. We particularly consider the implications of modeling choices about spatial relationships between polluters and residents, and about conditioning variables. Next, we evaluate the theory and evidence for four possible mechanisms that may lie behind these patterns: disproportionate siting on the firm side, “coming to the nuisance” on the household side, market-like coordination of the two, and discriminatory politics and/or enforcement. We argue that it is unclear how much weight each of these mechanisms carry in giving rise to the observed distribution, and that much previous research uses methodologies that

are inherently indecisive. Further research is needed, as we cannot hope to address injustices if we do not understand their origins. Finally, we offer an overview of some policy options before concluding.

Our discussion is focused almost exclusively on US topics, although similar questions have arisen in many parts of the world, including questions about land use, forest preservation, international disposal and management of toxic (including electronic) waste, and the disproportionate impacts of climate change. While there may be additional considerations with respect to the global or international nature of these cases, much of our discussion in the US context applies.

Modeling Exposure to Environmental Hazards

Early empirical work in environmental justice focused primarily on the demographics of people living near undesirable land uses such as hazardous waste sites and landfills.² Other papers have considered large air polluters, such as those listed in the US Toxic Release Inventory.³ However, proximity to nuisances may not capture actual risk exposure. Not every hazardous waste site handles an equal amount of waste, nor does every TRI site emit an equal amount of pollution. Moreover, pollutants vary in their toxicity. For example, beryllium (released from the burning of coal) is over three million times more hazardous as an air toxic than the same amount of dichlorotetrafluoroethane (often used as a refrigerant). Accordingly, some researchers have considered emission levels (for example, the weight of releases) and chemical toxicity rather than simply proximity to a site when defining exposure.⁴ Others have gone further by incorporating air quality models that characterize air pollutant dispersion across space, the better to capture the actual health hazards that populations face.⁵

Overall, while recent work has developed more nuanced and defensible measures of exposure, the overall finding that low-income households and people of color have greater exposure to environmental hazards is broadly supported by the application of these alternative measures. Moreover, the patterns seem persistent: following up on its seminal report of 1987 documenting environmental justice correlations, the United Church of Christ (Bullard, Mohai, Saha, and Wright

² Work along these lines includes Anderton, Anderson, Oakes, and Fraser (1994), Baden and Coursey (2002), Been (1997), Boer, Pastor, Sadd, and Snyder (1997), Cameron, Crawford, and McConnaha (2012), Currie (2011), Depro, Timmins, and O'Neil (2015), Gamper-Rabindran and Timmins (2011), Goldman and Fitton (1994), Persico, Figlio, and Roth (2016), Commission for Racial Justice, UCC (1987), and Bullard, Mohai, Saha, and Wright (2007).

³ Examples include Ringquist (1997), Sadd, Pastor, Boer, and Snyder (1999), and Wolverton (2009).

⁴ Examples include Arora and Cason (1999), Banzhaf and Walsh (2008, 2013), Bowen, Salling, Haynes, and Cyran (1995), Brooks and Sethi (1997), Kriesel, Centner, and Keeler (1996), and Ringquist (1997).

⁵ Examples include Ash and Fetter (2004); Depro and Timmins (2012); Depro, Timmins, and O'Neil (2015); Morello-Frosch and Jesdale (2006); and Morello-Frosch, Pastor, and Sadd (2001).

2007) found that disparities in race with respect to the location of hazardous waste remained largely unchanged 20 years later.

Statistical Controls: The Locus of Injustice

The most basic—and most robust—environmental justice pattern in the data is the simple correlation between pollution and poverty and/or people of color. But, in the context of linear regression, questions arise as to the importance of additional statistical controls. In an influential study that first questioned the robustness of some environmental justice research, Anderton, Anderson, Oakes, and Fraser (1994) found no evidence that race is significantly correlated with the location of waste facilities after including socioeconomic controls. Instead, they found that the most salient feature was the intensity of manufacturing employment, suggesting that the observed correlations may arise through local matching in the labor market, rather than matching pollution per se to residents. Similarly, a number of subsequent studies have questioned whether race is still significant after conditioning on income and proxies for wealth, or other characteristics of land and neighborhoods that might drive firms' production decisions. Others have criticized the methods used by Anderton et al. with respect to spatial relationships (Mohai and Saha 2006; Mohai, Pellow, and Roberts 2009), a point to which we return below.

Faced with these questions of robustness of the results, our view is that the literature has tended to dive too quickly into technical discussions of data aggregation and estimation, without pausing to consider the *question* being asked. If the question is about determining the social *causes* of environmental justice correlations, it may well make sense to include statistical controls. For example, Mohai and Bryant (1992) and Hamilton (1995) posit a number of reasons why racial groups may be directly correlated with exposure even after controlling for income and wealth, including taste-based discrimination by firms, racial discrimination in the housing market, and differential political clout and access to legal resources. A finding that race is not correlated with pollution after controlling for socioeconomic status might lead one to reject some of these hypotheses. Similarly, a study of firms' siting decisions (like Wolverton 2009) might want to control for the costs of land, labor, and transport, which are likely to affect a firm's profits.

On the other hand, it is not clear that it is important to distinguish race from class when establishing the *existence* of an environmental inequity. One could still argue that there is an injustice—even an injustice at the level of racial groups—when there are inequities in the simple correlations, even if these correlations are the result of socioeconomic processes. Simply because the inequity is mediated through some mechanism does not mean it isn't there.

Spatial Relationships

Until recently, environmental justice studies typically assumed that the population "exposed" to a nuisance coincides with those people living in the same geographic unit as the nuisance, such as a census tract or zip code—an approach known as "unit-hazard coincidence" (McMaster, Leitner, and Sheppard 1997; Mohai

and Saha 2006). The unit-hazard coincidence approach is straightforward, but is problematic for at least three reasons. First, it implicitly assumes that exposure to hazards is distributed equally within the geographic unit. However, geographic units like counties and census tracts vary greatly in size. For example, census tracts are drawn to have similar numbers of people, which means they are smaller in more densely populated urban areas. Unit-hazard coincidence thus assigns a smaller exposure area around facilities in more urban areas. Second, when nuisances are located near geographical boundaries, unit-hazard coincidence ignores exposures in adjacent areas that may be quite close by, while assigning exposure to parts of the coincident geographic unit that may be far from the nuisance. This introduces measurement error, which is likely to attenuate estimates of environmental justice correlations. Third, when geographic units like tracts are created to be fairly homogenous, the unit-hazard approach will mechanically extend local correlations to wider areas. For example, perhaps only a very local area near a nuisance is made up of one demographic group, but through the creation of homogeneous geographic units, that local area will be systematically combined with similar areas even if they are randomly distributed. This would tend to exaggerate any environmental justice correlations.

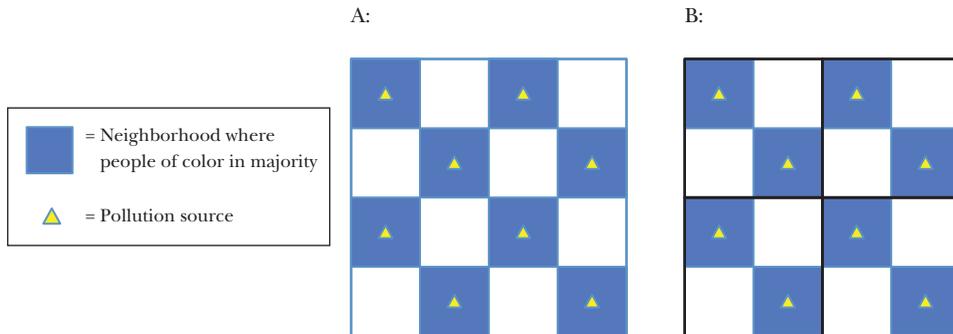
An alternative approach to measuring pollution exposure looks at the population within some distance of a site (or, alternatively, the nuisances or emissions near an arbitrarily drawn geographic unit). In the context of multiple regression and the question of sensitivity to included regressors discussed above, Mohai and Saha (2006) and Ringquist (2005) find that race is more likely to remain correlated with pollution when using distance-based measures rather than unit-hazard coincidence. Additionally, Mohai and Saha show that employing the unit-coincidence definition results in both larger tracts in the treatment group as well as tracts whose centroids are farther from facilities, compared to a distance-based measure. These patterns increase the likelihood of misclassifying exposure. This, along with the growing availability of new spatially resolute geocoded microdata and GIS technologies to analyze them, leads us to conclude that distance-based measures are superior to unit-coincidence.

Spatial Scale: The Ecological Fallacy

When measuring the correlation between pollution and demographics, the “ecological fallacy” can arise when inferring relationships between individual units (like households) from larger, more aggregated units (like counties) that contain those units. Some authors have raised concerns that the observed correlations between race and pollution found at the larger community level may potentially be subject to the ecological fallacy (for example, Anderton et al. 1994).

Figure 4 illustrates this concept. In Figure 4A, pollution sources (denoted by small triangles) are perfectly correlated with neighborhoods in which people of color are a majority (represented by larger, shaded squares), leading one to conclude that pollution is perfectly correlated with race. Suppose, however, that this relationship is estimated based on larger units of geography that aggregate neighborhoods according to the bold, dark lines in Figure 4B. Instead of 16 neighborhoods, the

Figure 4

The Ecological Fallacy

Note: Figure 4 illustrates the ecological fallacy. In panel A, pollution sources (denoted by small triangles) are perfectly correlated with neighborhoods in which people of color are a majority (represented by larger, shaded squares), leading one to conclude that pollution is perfectly correlated with race. Suppose, however, that this relationship is estimated based on larger units of geography that aggregate neighborhoods according to the bold, dark lines in panel B. Viewed at this more aggregated level, white neighborhoods and neighborhoods with people of color are now equally exposed to pollution sources, so one would conclude that there is no correlation between race and pollution.

data now is viewed as containing only four neighborhoods. Viewed at this more aggregated level, white neighborhoods and neighborhoods with people of color are now equally exposed to pollution sources, so one would conclude that there is no correlation between race and pollution.

The ecological fallacy teaches that the relationship estimated from aggregated data is only equal to the relationship at the micro level if there are no group-level effects correlated with pollution. This extreme assumption is not likely to hold. For example, if individuals have any peer preferences creating segregation between larger geographic units, or if the boundaries of geographic units are systematically gerrymandered in some way, we might expect exaggerated findings of correlations between pollution and demographics at a higher geographic level. If, on the other hand, communities follow a chessboard-like configuration due to clustering (as illustrated), then aggregation can mask inequitable exposure.

Baden, Noonan, and Turaga (2007) surveyed 110 environmental justice studies to assess the impact of the unit of analysis on their findings. They also conducted their own analysis to examine exposure to hazardous waste sites. Overall, they find that evidence of racial, ethnic, and income inequities becomes stronger when using smaller units of analysis (like tract and block group). Our earlier example contrasting Figures 2 and 3 illustrates these differences as well.

More recently, with increased data availability and advancements in computing, studies have employed individual-level data. In work that uses around 11 million births across five states between 1989 and 2003, Currie, Greenstone, and Moretti (2011) compare the characteristics of mothers in areas within two kilometers of

Superfund and TRI sites to those living farther. They find that black mothers are respectively 0.7 and 5.3 percentage points more likely to live near Superfund and Toxic Release Inventory sites; for Hispanics the respective values are 0 and 4.0. The results of many similar studies in recent years likewise echo the persistence of these environmental inequities (for example, Currie and Neidell 2005; Currie, Graff-Zivin, Meckel, Neidell, and Schlenker 2013; Persico, Figlio, and Roth 2016). It thus seems that the ecological fallacy tends to mask environmental injustices in more aggregate data, rather than the reverse.

Harms of Disparate Environmental Exposure

Ultimately, concerns about environmental justice come down to inequities in health outcomes. But does inequitable pollution exposure actually translate into inequitable outcomes? Plenty of indirect evidence suggests that it does. A large epidemiological literature has established a correlation between air pollution and human health, even conditioning on demographics like race and income (for a review, see Hoek et al. 2013). More recently, a wave of “quasi-experimental” studies has corroborated this finding and given it a stronger causal interpretation, especially for outcomes in young children, but also adults (Chay and Greenstone 2003; Currie and Neidell 2007; Currie and Walker 2011; Currie, Davis, Greenstone, and Walker 2015; Schlenker and Walker 2016). The results seem to hold for proximity to Superfund sites as well (Currie, Greenstone, and Moretti 2011; Persico, Figlio, and Roth 2016).

Given that the poor and minorities live closer to Superfund sites and to large air polluters, it stands to reason that they suffer more from the adverse consequences of such proximity. Consistent with this interpretation, Chay and Greenstone (2003) find that the county-level impact of large changes in Total Suspended Particulates on infant mortality for blacks is 1.6 times that for whites. If there are no racial differences in the biological responses to pollution, then these disparate health affects must come either from differences in other socioeconomic factors (like interactions with other burdens from poverty) or from uneven exposure to pollution within a county.

The effects on educational attainment and health of young children raise additional considerations about intergenerational equity. The “fetal origins” hypothesis (discussed in this journal by Almond and Currie 2011) posits a biological mechanism through which *in utero* health can persist through adulthood and affect long-term health, human capital accumulation, labor market outcomes, family structure, and welfare dependency (on this topic, see also Black, Devereux, and Salvanes 2007; Currie and Moretti 2007; Oreopoulos, Stabile, Walld, and Roos 2008; Royer 2009; Sanders 2012; Figlio, Guryan, Karbownik, and Roth 2014). Aizer, Currie, Simon, and Vivier (2018) and Persico, Figlio, and Roth (2016) find that exposure to lead and Superfund sites impacts test scores and other educational outcomes, and that reductions in such exposure substantially reduces the gap in educational outcomes between disadvantaged and other children. Thus, pollution today may have inequitable effects not only on today’s poor, but on the future poor as well. Indeed, it may contribute to perpetuating poverty traps.

Potential Mechanisms

Many of the studies documenting correlations among race, income, and pollution are unable to distinguish among alternative causal stories. Yet understanding the causal channels is important for two reasons. First, it helps to narrow down the locus of injustice. Is it based in actions by firms, either that are intentionally discriminatory or that have a discriminatory effect? In the underlying distribution of income? In differential patterns of participation by households in siting and permitting decisions? In enforcement decisions by governments? Second, understanding causal channels is crucial for designing policies meant to reverse the observed correlations.

Because modern econometrics has put such a heavy emphasis on causal identification, identifying the relative importance of these mechanisms may be an area where economists can best contribute to the environmental justice discussion. In this section, we consider four categories of causal mechanisms that could possibly give rise to environmental justice correlations: disproportionate siting by firms, “coming to the nuisance” on the household side, market-like coordination of the two in a process we will describe as Coasean bargaining, and discriminatory politics and/or enforcement.

Disproportionate Siting

A central focus of the environmental justice literature long has been whether, taking residential locational patterns as given, firms site (or historically have sited) polluting activity in poor neighborhoods and/or neighborhoods with people of color. Such disproportionate siting might occur for three broad reasons.

First, firms may be engaging in taste-based discrimination by incorporating into their decision-making a preference for protecting whites from pollution or indulging a malevolent preference to harm other groups (Becker 1957). Many economists, when first hearing about environmental justice concerns, assume this is what activists and researchers from other disciplines have in mind. In our experience, though, few activists or noneconomists think in these simplistic terms. Rather, they have a much more sophisticated understanding of the socioeconomic processes at work. Second, firms might site their polluting activity based on local economic conditions, which in turn are correlated in space with residential demographic patterns. For example, firms might seek access to inexpensive land, low-wage labor, or transportation networks (Wolverton 2009). These features might happen to be correlated with locations of poorer households for any number of reasons. Poorer households also might seek inexpensive land, for example, and they have lower wages almost by definition. Also, the correlations might arise indirectly from other types of discrimination. For example, industrial facilities may be attracted to locations near expressways or railroads, but those transportation routes might be located where they are because of past discriminatory transportation siting. Third, government agencies themselves make decisions that affect the location of such facilities, perhaps through the permitting process or other incentives that steer firms to such locations.

At the time of siting, firms do appear to go to areas that have a disproportionate share of people of color (Been 1997; Pastor, Sadd, and Hipp 2001; Baden and Coursey 2002). But modeling firm location as a decision variable, Wolverton (2009) finds that this pattern seems to arise more from economic factors such as land, labor, and access to transportation, rather than directly from local demographics.

Coming to the Nuisance

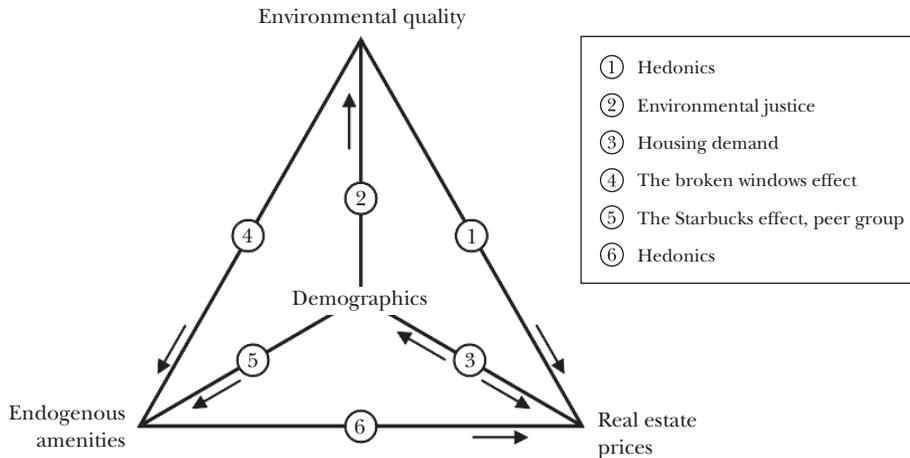
An alternative approach takes the pattern of pollution as given and considers the possibility that the households move based on their willingness to pay (and ability to pay) for a clean environment (Been 1994, 1997; Hamilton 1995). For economists, this perspective is an application of Tiebout's (1956) canonical model of residential sorting. In Tiebout's theory, households choose a location subject to a budget constraint, while taking into account desirable neighborhood amenities like a clean environment, green spaces, school quality, public safety, and access to employment centers and retail outlets. Because households prefer nicer neighborhoods, their demand for such neighborhoods is higher and, hence, *ceteris paribus*, so is the price of housing in those neighborhoods. Households therefore make tradeoffs between consumption and local neighborhood amenities, "voting with their feet" to reveal their willingness to pay more for public goods through higher gross-of-tax housing costs. How much a household is willing to pay for these amenities depends on its preferences and budget.

To imagine how this process works, consider a pair of neighboring communities with different levels of amenities and housing prices. A high-income household will obtain greater utility from the high-amenity community, and bid up prices there based on their willingness to pay. A lower-income household might also prefer the high-amenity community, but is not willing to pay the higher price, preferring to prioritize necessities like food and clothing. Essentially, it gets out-bid in the market for environmental quality. In this way, households "sort" by income across levels of amenities, a process also known as "stratification." Poorer households end up in more polluted areas, just as they obtain less of many of the other things money can buy. Moreover, if one demographic group (say, whites) is richer than another group throughout the distribution, then the poorer group will have more exposure to pollution on average (Banzhaf and Walsh 2013).

An initial sorting process like this may lead to additional effects on neighborhood characteristics as well, effects that could reinforce the initial sorting patterns. Figure 5 displays a pyramid of such relationships that could give rise to correlation between pollution and demographics in equilibrium (Banzhaf and McCormick 2012). Each vertex represents an outcome, and the lines connecting vertices represent relationships.

Line 2 depicts the direct relationship between environmental quality and demographics, documented in the environmental justice literature. Lines 1 and 3 show how this relationship can arise indirectly as areas with higher environmental quality have higher prices (as documented by the large "hedonics" literature), and richer people can afford higher prices as income effects shift out their demand.

Figure 5

Pyramid of Environmental Gentrification

Source: Banzhaf and McCormick (2012).

Meanwhile, if pollution signals neglect of, or disregard for, the neighborhood, polluting facilities may in this way undermine the provision of other local public goods, a kind of “broken windows” effect (Line 4). Additionally, once an initial sorting occurs because of the pollution, different demographic groups may create different neighborhood environments, based on their taste or their ability to pay for them. They may attract different types of retail (the so-called “Starbucks effect”), and richer neighborhoods may have a greater capacity than poorer neighborhoods to provide public safety, school quality, and so forth (Line 5). Such effects can further drive sorting, creating a multiplier effect (Banzhaf and Walsh 2013; O’Sullivan 2005; Sethi and Somanathan 2004). As Schelling (1969) showed many years ago, if people have even modest preferences to be with their own racial or ethnic group (homophily) for any reason, the dynamics of sorting can create “tipping patterns” such as white flight, driving further segregation. Such effects can also feed back on housing prices (Line 6).

To the extent that these Tiebout sorting processes by households are responsible for the observed correlations between demographics and pollution, there are four implications with some relevance for policy. First, the observed patterns may be “efficient”—*given the underlying distribution of resources*. That is not to say the outcome is “best” in a broad social welfare sense, but only to say that the poor may be doing the best they can with what they have. Second, to the extent that Tiebout sorting by households explains disparate exposure to pollution, it pushes back the locus of injustice from environmental inequities per se to the underlying distribution of income. In a similar vein, inequalities in the level of housing, energy consumption, and food consumption are all evidence of a deeper inequality of income and wealth. Third, if as illustrated in Line 5 of Figure 5, households sort based on homophily,

or have preferences for other amenities that are shared by households like themselves, it can create a kind of “tax” on the ability to obtain a clean environment. For example, people of color who want a clean environment not only have to pay in the form of higher housing costs, but also to live in a majority-white community, for which they might perceive little benefit (Banzhaf and Walsh 2013; Ford 1994).

Fourth, cleaning up locally polluted areas may increase the demand for local housing, bidding up housing prices (Line 1), a phenomenon known as “environmental gentrification” (Sieg, Smith, Banzhaf, and Walsh 2004; Banzhaf and McCormick 2012). For example, cleanup of Superfund sites (Gamper-Rabindran and Timmins 2013), brownfields (Haninger, Ma, and Timmings 2017), or hazardous waste sites more generally (Taylor, Phaneuf, and Li 2016) have been estimated to yield housing price increases of 5 to 15 percent. If high-income households bid up housing prices by their own willingness to pay, and if that willingness to pay exceeds that of the poor, the perverse result is that cleaning up pollution in a poor neighborhood can harm the poor, as prices increase by more than their willingness to pay. Homeowners, of course, also have a more valuable asset. But the poor are mostly renters, and renters only face higher rents, while landlords reap the gains. In this way, environmental justice considerations are not only wrapped up in the underlying distribution of income, but also potentially in historical policies such as “red-lining” (the practice of making mortgages less available to those living within certain neighborhoods) that have exacerbated disparities in homeownership (Aaronson, Hartley, and Mazumder 2017).

Empirical researchers have tried to examine the extent to which demographic factors in an area shift after either siting or closure/exit of polluting facilities, using difference-in-differences or similar methodologies. These studies have yielded mixed findings.⁶ Many of them have approached the problem as a chicken-or-the-egg question: Which came first, the siting of facilities in a poor, non-white neighborhood, or the sorting of such households near pollution? However, this problem has proven difficult to unscramble. There are at least four reasons why *changes* in demographics may not appear to be correlated with *changes* in pollution, even when cross-sectional correlations truly are driven by Tiebout sorting.

First, comparisons of the way different locations have evolved over time are complicated by general equilibrium effects in space. For example, imagine that pollution is cleaned up in a predominantly poor community, triggering in-migration. If the in-migrants are the poorest members of the comparison community, then *both* communities can become richer on average, making the prediction for differences-in-differences ambiguous, except for large changes (Banzhaf and Walsh 2008). The story is even more complicated for demographic groups (like racial

⁶ Evidence in favor of post-siting (or post-clean-up) sorting can be found in Baden and Coursey (2002), Banzhaf and Walsh (2008), Depro, Timmins, and O’Neil (2015), Gamper-Rabindran and Timmins (2011), Lambert and Boerner (1997), Oakes, Anderton, and Anderson (1996), and Wolverton (2009). Evidence against can be found in Been (1997), Cameron, Crawford, and McConaha (2012), Cameron and McConaha (2006), Eckerd (2011), Greenstone and Gallagher (2008), Messer et al. (2006), and Pastor, Sadd, and Hipp (2001).

groups), because the effects additionally depend on the relative density of each group at the marginal level of income where adjustments are occurring (Banzhaf and Walsh 2013).

Second, Depro, Timmins, and O’Neil (2015) show that there is an even more fundamental problem: without additional structure, individual sorting behavior with respect to pollution is not even identified from aggregate population changes. They construct examples illustrating that even when preferences for environmental quality versus other consumption leads people of color to differentially come to a nuisance, it does not necessarily follow that the percentage of residents who are of color will increase more in the polluted community than elsewhere. The problem comes from using aggregate data without observing the underlying substitution patterns implied by the specific population flows in the micro data.

Third, more generally, dynamic models often introduce the possibility of some hysteresis, or stickiness, in community compositions over time. For example, suppose an area initially is polluted, households sort on this pollution, and then the area is cleaned up. If perceptions are sticky, so that people believe the area is still dirty, or if some “stigma” has become associated with the original state (as in the broken windows effect), the initial sorting may not be reversed (Cameron and McConnaha 2006; Messer, Shultze, Hackett, Cameron, and McClelland 2006).

Finally, the multiplier effects discussed above also can create hysteresis. Changes in local amenities resulting from demographic sorting will themselves have feedbacks on housing prices, which induce further changes in sociodemographics. Consequently, even if the nuisance is eventually cleaned-up, housing prices may not bounce back because of these other changes, which themselves occurred as a result of Tiebout sorting (a sequence of Lines 5, 6, and 3 in Figure 5). Indeed, Banzhaf and Walsh (2013) show that if such effects become more salient after cleaning up pollution—because there is less reason to sort on environmental quality once it is cleaned up—then the reduced-form correlation between pollution and demographics can become *stronger* after cleanup, as racial differences increase and pollution differences decrease between communities. Consequently, scholars looking for such a reversal may fail to find it in the difference-in-differences.

Although the research literature speaking to coming-to-the-nuisance effects is large and growing, our view is that it still has not come to grips with many of these shortcomings. As a consequence, the jury is still out on the empirical relevance of this hypothesis. Depro, Timmins, and O’Neil (2015) offer one promising path forward—modeling substitution patterns and unobserved heterogeneity with a structural model in order to uncover heterogeneous willingness to pay for environmental patterns and, hence, sorting patterns. Future work might make more use of micro data describing individual moves, which would allow weaker modeling assumptions.

Coasean Bargaining

Firms have preferences over where to locate their industrial facilities and a willingness to pay to locate at a certain place. Households have a tolerance for pollution

and some willingness to accept compensation for industrial activity nearby. Whereas the previous two subsections take each of these perspectives in isolation, a Coasean perspective sees (the potential for) transactions between the two sides.

The Coase theorem holds that, under well-defined property rights and in the absence of transaction costs, it does not matter who holds the property rights to the use of the environment because negotiation and market transactions will ensure the same, efficient use of resources (Coase 1960). Through negotiation, the right to pollute (or to be spared pollution) will end up in the hands of the individuals or firms who value it most, and all parties will be appropriately compensated for any nuisance or foregone profits they consequently bear. However, the distribution of payments in this structure will depend on the initial allocation of property.

Consider a facility that emits pollution into the surrounding environment at zero private cost incurred by the facility. Each additional unit of pollution emitted creates a benefit to the facility in terms of not having to use an expensive abatement technology or not having to forego production. These marginal benefits to the facility from emitting pollution are declining and the facility will choose to generate additional units of pollution up to the point where marginal benefits of emitting pollution fall to zero. Suppose that local residents hold the relevant property rights, in the sense that they can veto polluting activities or accept them. Even if legislation does not explicitly recognize this right, local residents may be able to assert it through tort law, zoning laws, holding up permitting processes, political protest, and so forth. In the absence of any sort of compensation, local residents would prefer that the facility release no emissions. From a Coasean perspective, there is an opportunity for trade, in which residents of the community agree to some level of pollution in exchange for compensation. Such payments may be cash transfers, or they may take the form of local jobs, investments in parks and community centers, and so forth.

In this situation, the exact amount of the payment the firm needs to make would be determined in some fashion between the two parties, according to their relative bargaining power. One aspect of this negotiation is that polluters have an incentive to locate where the local residents are willing to accept relatively low compensation to offset accepting a degree of pollution; also, a community that requires relatively low compensation for accepting pollution will be compatible with a higher efficient level of pollution. For example, their willingness-to-accept compensation for pollution will tend to be lower in remote locations with fewer people, so that total injuries are lower. Alternatively, those with lower income levels may also be willing to accept lower levels of compensation for injuries. In either case, Coasean bargaining theory would treat this incentive as one leading to economic efficiency. However, this latter incentive might also give rise to environmental justice correlations (Hamilton 1995). It is likely to be poorer households who have a lower willingness to accept compensation, perhaps because they have a high marginal utility of income and prioritize consumption of other important goods.

As with the Tiebout mechanism, the model of Coasean bargaining suggests that the observed distribution of pollution is efficient—given the distribution of income. Likewise, it also shifts the locus of injustice back to inequality in the

underlying distribution of income and wealth. However, in the Coasean case, the environmental resource is a valuable asset, and its allocation represents part of the distribution of wealth.

Combining Coasean insights with the insights of the environmental justice movement, there would be important justice arguments for allocating the right to the environment to local communities. In this case, local communities could then keep the right to be free of polluting facilities, or negotiate with polluters, as they saw fit. If such communities have full information and full power to bargain (admittedly big “ifs”), they cannot be worse off if they accept compensation and allow polluting facilities to operate. This observation highlights one potential area of tension in the understanding of environmental justice. While in many cases, a focus on equity of exposure to pollution (a kind of distributive justice) and a focus on the ability to participate in decision-making (procedural justice) run together, they do not always. Full environmental justice entails sovereignty over environmental decisions, and thus the right to accept polluting firms as well as to reject them (Foster 1998).

A Coasean bargaining scenario raises the obvious question of why, then, there is an environmental justice movement protesting siting that leads to local pollution. Presumably, disadvantaged groups are bearing the brunt of pollution exposure but do not feel they are receiving the compensating benefits of such Coasean bargains. Why not? One possibility, of course, is that local communities do not have the property rights to the environment, so that firms can locate where they wish, perhaps with discriminatory effect. Allowing for more nuance, it may be that property rights are ambiguous and are left open to being claimed through political action, or being exploited through channels such as zoning and permitting processes. To some extent, the environmental justice movement might be interpreted as an effort to claim rights. Alternatively, it might be interpreted as a protest of the extra difficulty environmental justice communities have to go through in claiming such rights. Such communities may be limited by, for example, less access to the corridors of power, less formal education, language barriers, and other such disadvantages (Hamilton 1995). In addition, communities facing environmental justice concerns may have difficulties overcoming the free rider problem on their side of the negotiation: whereas the benefits of polluting are a concentrated interest for one firm, the costs of polluting are dispersed among all residents of a jurisdiction. As a consequence, communities facing environmental justice concerns, when in negotiations, may end up systematically with a small share of the Coasean surplus. Indeed, firms might systematically aim to locate in communities which will be in a weaker bargaining position.

Stepping back, we might ask what it even means to allocate rights to local communities, and who speaks for the community in negotiations. Whether the actual decisionmakers at the local level are local government officials or community organizations, it will commonly be true that those bargaining on behalf of victims are not actually bearing the costs of the pollution—and thus the negotiation may lead to more pollution and less compensation than if the victims bargained for themselves.

Such themes seem to have been illustrated by the case of Kettleman City, California, described in Cole and Foster (2001). In 1988, a waste management firm proposed building a toxic waste incinerator at a nearby dump site. Located in the San Joaquin Valley, Kettleman City was 90 percent Hispanic, with 40 percent of residents speaking no English. Through inadequate provision of public notice, the begrudging provision of translators at public meetings, and the scheduling of those meetings in difficult-to-reach locations, it was clear that information asymmetries were part of a strategy to inhibit local participation. Despite vigorous protests from the residents of Kettleman City, Kings County initially approved the deal. The county was to receive \$7 million annually from the deal, but these benefits were spread over a 1,400 square mile rural county, with a very different demographic, while the environmental injuries were concentrated in Kettleman City.

The Kettleman City example demonstrates how political economy problems can overcome Coasean forces even at the county level. Such forces may be even stronger at the state level. The Cerrell Report (Cerrell Associates 1984), a consulting report requested by the state of California Waste Management Board, provides an infamous example of an effort to direct the siting of polluting facilities towards communities that would be ineffective bargainers. The report identified characteristics of local communities that would not protest the location of waste sites in their area—in particular, people without a college education, the poor, Catholics, and those “not involved in voluntary associations.”

More systematically, Timmins, and Vissing (2017) examine the content of leases signed between shale gas operators and households in Tarrant County, Texas, for the rights to extract natural gas. The terms of these leases dictate both payments in the form of royalty compensation and protective clauses designed to reduce health and environmental risks from the extraction process. After conditioning on income, Vissing and Timmins find that race and, interestingly, English-speaking are correlated with lease terms (like protective clauses) and royalty compensation.

There are other cases in which we see some evidence of Coasean logic at work, in the sense of the comparative analysis of transactions costs. For example, Coasean compensation appears to take place in the form of host fees collected by neighborhoods near landfills. Jenkins, Maguire, and Morgan (2004) find that citizen participation in host fee negotiations leads to greater host compensation. Similarly, Hamilton (1993, 1995) finds that communities better able to organize politically (as proxied by higher voter turnout) are less likely to see local firms expand their processing of hazardous wastes (see also Brooks and Sethi 1997; Arora and Cason 1999).

Political Economy and Government

Governments can affect the distribution of pollution in a number of ways, including through legislation, bureaucratic monitoring and enforcement patterns, and court enforcement patterns. Regulators must choose how to allocate the policy tools at their disposal, prioritizing regulation and remediation across various sites in the face of resource and time constraints. They may make decisions based on technical factors (like size of operating facility, hazardousness of processed materials,

potential risks to surrounding neighborhood), on polluter factors (like polluter's ability to pay, polluter's violation history, polluter's negotiation/bargaining power), and on the weight they give to interest groups that may "capture" the regulatory process. This raises the possibility that regulators could be a source of inequitable exposure to environmental nuisances. Households with the highest willingness to pay for avoiding pollution, combined with the greatest ease and ability to influence government, may exert the most pressure on government agencies (Becker 1983). Thus, as with Coasean bargaining, differences in the ability to organize, be heard, or be pivotal to government officials can drive different degrees of influence.

In an early study, Lavelle and Coyle (1992) concluded that polluters accused of violating environmental regulations faced lower enforcement penalties if they were in areas with more people of color. They also found that, in such areas, cleanup times were longer and cleanup solutions were less stringent. Viscusi and Hamilton (1999) re-examined remediation activities, focusing on the choice of post-cleanup standards. They found, perhaps surprisingly, that regulators impose *stricter* risk targets in areas with more people of color. The same is also true (less surprisingly) in areas with greater potential for collective action. Income and proportion non-white do not appear to affect the cost per cancer case avoided, but voter turnout (a proxy for collective action) does. More recently, Gray and Shadbegian (2004) and Shadbegian and Gray (2012) have studied the determinants of regulatory stringency in communities near polluting facilities, with a focus on the application of penalties and the frequency of inspections. They again find that measures of the potential for collective action are important determinants of enforcement activities, but also that race does not have an independent effect and that the effect of income is mixed.

We conclude that there is evidence that regulatory actions are at least correlated with the political power of local communities. Interestingly, this finding is similar to that used in the literature to evaluate the extent of Coasean processes. This similarity highlights the important connection between Coasean processes and political economy, as both are tied up in property rights and the enforcement of those rights.

Discussion and Conclusion

In 1993, some ten years after it was sited, the Warren County landfill in North Carolina was found to be leaking PCBs. Eventually, 81,500 tons of contaminated soil were excavated and burned at a cost of \$18 million, seemingly justifying the initial protests over the decision. Moreover, as the more recent case of lead contamination in the water of Flint, Michigan, indicates, controversies over environmental justice are still with us. Like Warren County, Flint is disproportionately poor and African American. And also like Warren County, it was higher levels of government that made decisions affecting local populations, populations that felt they did not have an adequate voice.

Cases like these highlight a first "no-brainer" policy response to environmental justice concerns: giving local populations a seat at the table when making decisions

that affect the local environment. In some cases, this may mean devolving decisions to more-local governments. In others, it may mean incorporating local comments into state-level or national regulatory reviews. As noted previously, the policy of the Environmental Protection Agency calls for “meaningful involvement” of all people.

Having a seat at the table can help environmental justice communities in other contexts as well. In particular, if outcomes are the result of Coasean bargains, then policies should be structured to assure that disadvantaged communities can bargain more effectively based on their preferences: for example, access to legal expertise and pollution disclosure policies can increase access to information about hazards and legal remedies. Here, environmental justice advocates can play an important role, by providing services that effectively reduce transactions costs for local residents and that level the playing field. Having such a voice might also minimize gentrification effects, if local residents propose new uses for previously contaminated lands that better fit the existing character of the community (NEJAC 2006).

Second, insofar as disproportionate exposure arises out of firm choices to locate in areas with people of color (for whatever reason), then such patterns affect the distribution of real wealth. Accordingly, environmental policies that target pollution directly may have a “double dividend” in being progressive as well as (potentially) efficiency enhancing, as recently argued by Bento, Freedman, and Lang (2015).

Third, people-based investments that target income inequality may be more fruitful than targeting environmental correlations, especially if sorting is the predominant force underlying environmental justice correlations. Such sorting would place the ultimate source of the correlations in the income distribution; it also implies that attempts to reverse environmental justice correlations may be accompanied by gentrification effects. Of course, this suggestion begs the question of how to reduce inequality. In a narrow sense, one might focus on ways for low-income renters and owners to have a wider range of affordable housing options, essentially giving them a property right in environmental improvements. As a result, they would be less likely to face a tradeoff between lower housing costs and exposure to pollution. One can also imagine a range of other compensatory benefits, like greater support for health care for pregnant women and programs to support the development of newborns and very young children. In general, the fact of disparate environmental effects on those with low socioeconomic status strengthens the arguments for redistribution to these groups in other forms.

The last two points raise a more general issue about the evaluation across income groups of policies that create environmental benefits and costs. Given the current distribution of pollution exposure, the direct effects of environmental improvements will generally be progressive in the sense that the improved quality of life and health should be enjoyed especially by those of lower socioeconomic status. But on the other side, the indirect effects of environmental improvement on housing prices (gentrification) and energy prices may be especially burdensome to the poor. Both effects could be better incorporated into regulatory decision-making. Current practice of the Environmental Protection Agency is to incorporate environmental justice into rulemaking; focus (though not exclusively) on assessing

the distribution of health risks and benefits; and use a constant willingness-to-pay to aggregate the benefits for all demographic groups (US EPA 2014, 2016).

However, when policies have differential costs as well as benefits, such a practice can distort benefit–cost tests. Indeed, in theory, a policy that harms everybody could pass benefit–cost tests with this practice. Suppose, for example, that the rich are willing to pay \$8 million per statistical life saved, and the poor \$4 million (because, having greater unmet needs, their opportunity cost of money is higher). Suppose there are equal numbers of rich and poor. Suppose finally that a policy would cost the poor \$5 million per statistical life saved (in higher prices) and cost the rich a very small amount but save no lives among the rich. Using an average value of life of \$6 million, the policy would pass a benefit–cost test. But everybody is made worse off by the policy. Given a concern that heterogeneous willingness-to-pay may favor higher-income groups, one could combine heterogeneous willingness-to-pay with distributional weights based on a social welfare function while making more effort to include indirect, general-equilibrium effects. This could give regulators a systematic way to implement environmental justice concerns while still giving people sovereignty in the sense of respecting their individual preferences (Adler 2016; Banzhaf 2011).

Key to this policy discussion is that any specific prescription is contingent upon *how* inequities arise. Tackling the remaining uncertainty about the relative importance of such causes is critical if we hope to address environmental injustice at a fundamental level. Armed with models of residential location, firm entry/siting decisions, and government decision-making, economists today are in a unique position to contribute to the discussion of “why” environmental injustice arises and to devise appropriate policy solutions.

■ *We thank the editors, especially Timothy Taylor, for helpful comments and suggestions.*

References

- Aaronson, Daniel, Daniel A. Hartley, and Bhashkar Mazumder. 2017. “The Effects of the 1930s HOLC ‘Redlining’ Maps.” Federal Reserve Bank of Chicago Working Paper no. 2017-12.
- Adler, Matthew D. 2016. “Benefit–Cost Analysis and Distributional Weights: An Overview.” *Review of Environmental Economics and Policy* 10(2): 264–85.
- Aizer, Aizer, Janet Currie, Peter Simon, and Patrick Vivier. 2018. “Do Low Levels of Blood Lead Reduce Children’s Future Test Scores?” *American Economic Journal: Applied Economics* 10(1): 307–41.
- Almond, Douglas, and Janet Currie. 2011. “Killing Me Softly: The Fetal Origins Hypothesis.” *Journal of Economic Perspectives* 25(3): 153–72.
- Anderton, Douglas L., Andy B. Anderson, John Michael Oakes, and Michael R. Fraser. 1994. “Environmental Equity: The Demographics of Dumping.” *Demography* 31(2): 229–48.
- Arora, Seema, and Timothy N. Cason. 1999. “Do Community Characteristics Influence Environmental Outcomes? Evidence from the Toxics Release Inventory.” *Southern Economic Journal* 65(4):

691–716.

Ash, Michael, and T. Robert Fetter. 2004. “Who Lives on the Wrong Side of the Environmental Tracks? Evidence from the EPA’s Risk-Screening Environmental Indicators Model.” *Social Sciences Quarterly* 85(2): 441–62.

Baden, Brett M., and Don L. Coursey. 2002. “The Locality of Waste Within the City of Chicago: A Demographic, Social, and Economic Analysis.” *Resource and Energy Economics* 24(1–2): 53–93.

Baden, Brett M., Douglas S. Noonan, and Rama M. R. Turaga. 2007. “Scales of Justice: Is There a Geographic Bias in Environmental Equity Analysis?” *Journal of Environmental Planning and Management* 50(2): 163–85.

Banzhaf, H. Spencer. 2011. “Regulatory Impact Analyses of Environmental Justice Effects.” *Journal of Land Use and Environmental Law* 27(1): 1–30.

Banzhaf, H. Spencer, ed. 2012. *The Political Economy of Environmental Justice*. Stanford University Press.

Banzhaf, H. Spencer, and Eleanor McCormick. 2012. “Moving Beyond Cleanup: Identifying the Crucibles of Environmental Gentrification.” Chap. 2 in *The Political Economy of Environmental Justice*, edited by H. Spencer Banzhaf. Stanford University Press.

Banzhaf, H. Spencer, and Randall P. Walsh. 2008. “Do People Vote with Their Feet? An Empirical Test of Tiebout’s Mechanism.” *American Economic Review* 98(3): 843–63.

Banzhaf, H. Spencer, and Randall P. Walsh. 2013. “Segregation and Tiebout Sorting: The Link Between Place-based Investments and Neighborhood Tipping.” *Journal of Urban Economics* 74: 83–98.

Becker, Gary S. 1957. *The Economics of Discrimination*. University of Chicago Press.

Becker, Gary S. 1983. “A Theory of Competition among Pressure Groups for Political Influence.” *Quarterly Journal of Economics* 98(3): 371–400.

Been, Vicki. 1994. “Locally Undesirable Land Uses in Minority Neighborhoods: Disproportionate Siting or Market Dynamics?” *Yale Law Journal* 103(6): 1383–1422.

Been, Vicki. 1997. “Coming to the Nuisance or Going to the Barrios? A Longitudinal Analysis of Environmental Justice Claims.” With Francis Gupta. *Ecology Law Quarterly* 24(1): 1–56.

Bento, Antonio, Matthew Freedman, and Corey Lang. 2015. “Who Benefits from Environmental Regulation? Evidence from the Clean Air Act Amendments.” *Review of Economics and Statistics* 97(3): 610–22.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2007. “From the Cradle to the Labor Market? The Effect of Birth Weight on Adult

Outcomes.” *Quarterly Journal of Economics* 122(1): 409–39.

Boer, J. Tom, Manuel Pastor Jr., James L. Sadd, and Lori D. Snyder. 1997. “Is There Environmental Racism? The Demographics of Hazardous Waste in Los Angeles County.” *Social Science Quarterly* 78(4): 793–810.

Bowen, William. 2002. “An Analytical Review of Environmental Justice Research: What Do We Really Know?” *Environmental Management* 29(1): 3–15.

Bowen, William M., Mark J. Salling, Kingsley E. Haynes, and Ellen J. Cyran. 1995. “Toward Environmental Justice: Spatial Equity in Ohio and Cleveland.” *Annals of the Association of American Geographers* 85(4): 641–63.

Brooks, Nancy, and Rajiv Sethi. 1997. “The Distribution of Pollution: Community Characteristics and Exposure to Air Toxics.” *Journal of Environmental Economics and Management* 32(2): 233–50.

Bullard, Robert D. 1983. “Solid Waste Sites and the Black Houston Community.” *Sociological Inquiry* 53(2–3): 273–88.

Bullard, Robert D. 1994. *Dumping in Dixie: Race, Class, and Environmental Quality*, 2nd edition. Westview Press.

Bullard, Robert, Paul Mohai, Robin Saha, and Beverly Wright. 2007. *Toxic Wastes and Race at Twenty 1987-2007: A Report Prepared for the United Church of Christ Justice & Witness Ministries*. United Church of Christ (UCC). http://www.ucc.org/environmental-ministries_toxic-waste-20.

Cameron, Trudy Ann, Graham D. Crawford, and Ian T. McConnaha. 2012. “Superfund Taint and Neighborhood Change: Ethnicity, Age Distributions, Household Structure.” In *The Political Economy of Environmental Justice*, edited by H. Spencer Banzhaf, 137–69. Stanford University Press.

Cameron, Trudy Ann, and Ian T. McConnaha. 2006. “Evidence of Environmental Migration.” *Land Economics* 82(2): 273–90.

Cerrell Associates. 1984. *Political Difficulties Facing Waste-to-Energy Conversion Plant Siting*. Prepared by Cerrell Associates, Inc. for California Waste Management Board.

Chay, Kenneth Y., and Michael Greenstone. 2003. “The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession.” *Quarterly Journal of Economics* 118(3): 1121–67.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. “Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility.” *American Economic Review* 104(5): 141–47.

- Coase, R. H.** 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3: 1–44.
- Cole, Luke W., and Sheila R. Foster.** 2001. *From the Ground Up: Environmental Racism and the Rise of the Environmental Justice Movement*. NYU Press.
- Commission for Racial Justice, United Church of Christ (UCC).** 1987. *Toxic Wastes and Race in the United States: A National Report on the Racial and Socio-Economic Characteristics of Communities with Hazardous Waste Sites*. <http://www.ucc.org/about-us/archives/pdfs/toxwrace87.pdf>.
- Currie, Janet.** 2011. "Inequality at Birth: Some Causes and Consequences." *American Economic Review* 101(3): 1–22.
- Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker.** 2015. "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings." *American Economic Review* 105(2): 678–709.
- Currie, Janet, Joshua Graff-Zivin, Katherine Meckel, Matthew Neidell, and Wolfram Schlenker.** 2013. "Something in the Water: Contaminated Drinking Water and Infant Health." *Canadian Journal of Economics* 46(3): 791–810.
- Currie, Janet, Michael Greenstone, and Enrico Moretti.** 2011. "Superfund Cleanups and Infant Health." *American Economic Review* 101(3): 435–41.
- Currie, Janet, and Enrico Moretti.** 2007. "Biology as Destiny? Short-and Long-Run Determinants of Intergenerational Transmission of Birth Weight." *Journal of Labor Economics* 25(2): 231–64.
- Currie, Janet, and Matthew Neidell.** 2005. "Air Pollution and Infant Health: What Can We Learn From California's Recent Experience?" *Quarterly Journal of Economics* 120(3): 1003–1030.
- Currie, Janet, and Matthew Neidell.** 2007. "Getting Inside the 'Black Box' of Head Start Quality: What Matters and What Doesn't." *Economics of Education Review* 26(1): 83–99.
- Currie, Janet, and Reed Walker.** 2011. "Traffic Congestion and Infant Health: Evidence from E-ZPass." *American Economic Journal: Applied Economics* 3(1): 65–90.
- Depro, Brooks, and Christopher Timmins.** 2012. "Residential Mobility and Ozone Exposure: Challenges for Environmental Justice Policy." Chap. 5 in *Political Economy of Environmental Justice*, edited by H. Spencer Banzhaf. Stanford University Press.
- Depro, Brooks, Christopher Timmins, and Maggie O'Neil.** 2015. "White Flight and Coming to the Nuisance: Can Residential Mobility Explain Environmental Injustice?" *Journal of the Association of Environmental and Resource Economists* 2(3): 439–68.
- Eckerd, Adam.** 2011. "Cleaning Up Without Clearing Out? A Spatial Assessment of Environmental Gentrification." *Urban Affairs Review* 47(1): 31–59.
- Figlio, David, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth.** 2014. "The Effects of Poor Neonatal Health on Children's Cognitive Development." *American Economic Review* 104(12): 3921–55.
- Ford, Richard Thompson.** 1994. "The Boundaries of Race: Political Geography in Legal Analysis." *Harvard Law Review* 107(8): 1841–1921.
- Foster, Sheila.** 1998. "Justice from the Ground Up: Distributive Inequities, Grassroots Resistance, and the Transformative Politics of the Environmental Justice Movement." *California Law Review* 86(4): 775–841.
- Gamper-Rabindran, Shanti, and Christopher Timmins.** 2011. "Hazardous Waste Cleanup, Neighborhood Gentrification, and Environmental Justice: Evidence from Restricted Access Census Block Data." *American Economic Review Papers and Proceedings* 101(3): 620–24.
- Gamper-Rabindran, Shanti, and Christopher Timmins.** 2013. "Does Cleanup of Hazardous Waste Sites Raise Housing Values? Evidence of Spatially Localized Benefits." *Journal of Environmental Economics and Management* 65(3): 345–60.
- Goldman, Benjamin A., and Laura J. Fitton.** 1994. *Toxic Wastes and Race Revisited: An Update of the 1987 Report on the Racial and Socioeconomic Characteristics of Communities with Hazardous Waste Sites*. Washington, DC: Center for Policy Alternatives.
- Gray, Wayne B., and Ronald J. Shadbegian.** 2004. "'Optimal' Pollution Abatement—Whose Benefits Matter and How Much." *Journal of Environmental Economics and Management* 47(3): 510–34.
- Greenstone, Michael, and Justin Gallagher.** 2008. "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program." *Quarterly Journal of Economics* 123(3): 951–1003.
- Hamilton, James T.** 1993. "Politics and Social Costs: Estimating the Impact of Collective Action on Hazardous Waste Facilities." *RAND Journal of Economics* 24(1): 101–125.
- Hamilton, James T.** 1995. "Testing for Environmental Racism: Prejudice, Profits, Political Power?" *Journal of Policy Analysis and Management* 14(1): 107–32.
- Hampson, Chris.** 2010. "Warren County and Environmental Justice: A Community Fighting Back." Senior Thesis, University of North Carolina at Asheville.
- Haninger, Kevin, Lala Ma, and Christopher Timmins.** 2017. "The Value of Brownfield Remediation." *Journal of the Association of Environmental and Resource Economists* 4(1): 197–241.
- Hoek, Gerald, Ranjini M. Krishnan, Rob Beelen, Annette Peters, Bart Ostro, Bert Brunekreef, and**

- Joel D. Kaufman. 2013. "Long-Term Air Pollution Exposure and Cardio-Respiratory Mortality: A Review." *Environmental Health* 12(43).
- Jenkins, Robin R., Kelly B. Maguire, and Cynthia L. Morgan. 2004. "Host Community Compensation and Municipal Solid Waste Landfills." *Land Economics* 80(4): 513–28.
- Kriesel, Warren, Terence J. Centner, and Andrew G. Keeler. 1996. "Neighborhood Exposure to Toxic Releases: Are There Racial Inequities?" *Growth and Change* 27(4): 479–99.
- Lambert, Thomas, and Christopher Boerner. 1997. "Environmental Inequity: Economic Causes, Economic Solutions." *Yale Journal on Regulation* 14(1): 195–234.
- Lavelle, Marianne, and Marcia Coyle. 1992. "Unequal Protection: The Racial Divide in Environmental Law." *National Law Journal* 15(3): S1–S12.
- McMaster, Robert B., Helga Leitner, and Eric Sheppard. 1997. "GIS-based Environmental Equity and Risk Assessment: Methodological Problems and Prospects." *Cartography and Geographic Information Systems* 24(3): 172–89.
- Messer, Kent D., William D. Schulze, Katherine F. Hackett, Trudy A. Cameron, and Gary H. McClelland. 2006. "Can Stigma Explain Large Property Value Losses? The Psychology and Economics of Superfund." *Environmental and Resource Economics* 33(3): 299–324.
- Mohai, Paul, and Bunyan Bryant. 1992. "Environmental Injustice: Weighing Race and Class as Factors in the Distribution of Environmental Hazards." *University of Colorado Law Review* 63: 921–32.
- Mohai, Paul, David Pellow, and J. Timmons Roberts. 2009. "Environmental Justice." *Annual Review of Environment and Resources* 34: 405–30.
- Mohai, Paul, and Robin Saha. 2006. "Reassessing Racial and Socioeconomic Disparities in Environmental Justice Research." *Demography* 43(2): 383–99.
- Morello-Frosch, Rachel, and Bill M. Jesdale. 2006. "Separate and Unequal: Residential Segregation and Estimated Cancer Risks Associated with Ambient Air Toxics in U.S. Metropolitan Areas." *Environmental Health Perspectives* 114(3): 386–93.
- Morello-Frosch, Rachel, Manuel Pastor, and James Sadd. 2001. "Environmental Justice and Southern California's 'Riskscape': The Distribution of Air Toxics Exposures and Health Risks among Diverse Communities." *Urban Affairs Review* 36(4): 551–78.
- National Environmental Justice Advisory Council (NEJAC). 2006. *Unintended Impacts of Redevelopment and Revitalization Efforts in Five Environmental Justice Communities*. Final Report.
- Noonan, Douglas S. 2008. "Evidence of Environmental Justice: A Critical Perspective on the Practice of EJ Research and Lessons for Policy Design." *Social Science Quarterly* 89(5): 1154–74.
- Oakes John Michael, Douglas L. Anderton, and Andy B. Anderson. 1996. "A Longitudinal Analysis of Environmental Equity in Communities with Hazardous Waste Facilities." *Social Science Research* 25(2): 125–48.
- O'Sullivan, Arthur. 2005. "Gentrification and Crime." *Journal of Urban Economics* 57(1): 73–85.
- Oreopoulos, Philip, Mark Stable, Randy Walld, and Leslie L. Roos. 2008. "Short-, Medium-, and Long-term Consequences of Poor Infant Health, An Analysis Using Siblings and Twins." *Journal of Human Resources* 43(1): 88–138.
- Pastor, Manuel, Jim Sadd, and John Hipp. 2001. "Which Came First? Toxic Facilities, Minority Move-in, and Environmental Justice." *Journal of Urban Affairs* 23(1): 1–21.
- Persico, Claudia, David Figlio, and Jeffrey Roth. 2016. "Inequality Before Birth: The Developmental Consequences of Environmental Toxicants." NBER Working Paper 22263
- Piketty, Thomas. 2014. *Capital in the Twenty-first Century*. Harvard University Press.
- Ringquist, Evan. 1997. "Equity and Distribution of Environmental Risk: The Case of TRI Facilities." *Social Science Quarterly* 78(4): 811–29.
- Ringquist, Evan J. 2003. "Environmental Justice: Normative Concerns, Empirical Evidence, and Government Action." Chap. 11 in *Environmental Policy: New Directions for the Twenty-First Century*, 5th, edition, edited by Norman J. Vig and Michael E. Kraft. Washington, DC: CQ Press.
- Ringquist, Evan J. 2005. "Assessing Evidence of Environmental Inequities: A Meta-Analysis." *Journal of Policy Analysis and Management* 24(2): 223–47.
- Sadd, James L., Manuel Pastor, J. Thomas Boer, and Lori D. Snyder. 1999. "'Every Breath You Take...': The Demographics of Toxic Air Releases in Southern California." *Economic Development Quarterly* 13(2): 107–123.
- Sanders, Nicholas J. 2012. "What Doesn't Kill You Makes You Weaker: Prenatal Pollution Exposure and Educational Outcomes." *Journal of Human Resources* 47(3): 826–50.
- Schelling, Thomas C. 1969. "Models of Segregation." *American Economic Review* 59(2): 488–93.
- Schlenker, Wolfram, and W. Reed Walker. 2016. "Airports, Air Pollution, and Contemporaneous Health." *Review of Economic Studies* 83(2): 768–809.
- Sethi, Rajiv, and Rohini Somanathan. 2004. "Inequality and Segregation." *Journal of Political Economy* 112(6): 1296–1321.
- Shadbegian, Ronald J., and Wayne B. Gray. 2012. "Spatial Patterns in Regulatory Enforcement:

Local Tests of Environmental Justice.” Chap. 9 in *The Political Economy of Environmental Justice*, edited by H. Spencer Banzhaf. Stanford University Press.

Sieg, Holger, V. Kerry Smith, H. Spencer Banzhaf, and Randy Walsh. 2004. “Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods.” *International Economic Review* 45(4): 1047–77.

Taylor, Laura O., Daniel J. Phaneuf, and Xiangping Li. 2016. “Disentangling Property Value Impacts of Environmental Contamination from Locally Undesirable Land Uses: Implications for Measuring Post-Cleanup Stigma.” *Journal of Urban Economics* 93: 85–98.

Tiebout, Charles M. 1956. “A Pure Theory of Local Expenditures.” *Journal of Political Economy* 64(5): 416–24.

Timmins, Christopher, and Ashley Vissing. 2017. “Environmental Justice and Coasian Bargaining: The Role of Race and Income in Lease Negotiations for Shale Gas.” <https://sites.duke.edu/christophertimmins/files/2017/02/>

[environmental-justice-coasian.pdf](#).

US EPA. 2014. *Guidelines for Preparing Economic Analyses*. US Environmental Protection Agency. Washington, DC <https://www.epa.gov/sites/production/files/2017-08/documents/ee-0568-50.pdf>.

US EPA. 2016. *Technical Guidance for Assessing Environmental Justice in Regulatory Analysis*. US Environmental Protection Agency. Washington, DC.

US General Accounting Office (US GAO). 1983. *Siting of Hazardous Waste Land-Fills and their Correlation with Racial and Economic Status of Surrounding Communities*. Report. <http://archive.gao.gov/d48t13/121648.pdf>.

Viscusi, W. Kip, and James T. Hamilton. 1999. “Are Risk Regulators Rational? Evidence from Hazardous Waste Cleanup Decisions.” *American Economic Review* 89(4): 1010–27.

Wolverton, Ann. 2009. “Effects of Socio-Economic and Input-Related Factors on Polluting Plants’ Location Decisions.” *BE Journal of Economic Analysis & Policy: Advances* 9(1): Article 14.

Economists (and Economics) in Tech Companies

Susan Athey and Michael Luca

PhD economists have started to play an increasingly central role in tech companies, tackling problems such as platform design, pricing, and policy. Major companies, including Amazon, eBay, Google, Microsoft, Facebook, Airbnb, and Uber, have large teams of PhD economists working to engineer better design choices. For example, led by Pat Bajari, Amazon has hired more than 150 PhD economists in the past five years, making it the largest employer of tech economists. In fact, Amazon now has several times more full-time economists than the largest academic economics department, and continues to grow at a rapid pace. Companies such as Coursera, Expedia, Microsoft, Netflix, Pandora, Uber, Yelp, and Zillow have also hired economists. Table 1 shows a list of some technology companies that have hired PhD economists, although the list is not comprehensive.

Hiring of PhD economists has happened at all levels, from newly minted PhDs heading directly to the tech sector up through chief economists plucked from tenured positions at prestigious academic departments. The types of positions also vary greatly. Much of the recent growth has focused on economists working directly on business problems, with only a small fraction of the work resulting in academic papers. In contrast, some companies, such as Microsoft, have a chief economist who manages teams focused directly on business problems, but also have many economists working out of research centers, publishing self-guided research in academic

■ *Susan Athey is Economics of Technology Professor, Stanford Graduate School of Business, Stanford, California. Michael Luca is Lee J. Styslinger III Associate Professor of Business Administration, Harvard Business School, Boston, Massachusetts. Their email addresses are athey@susanathey.com and mluca@hbs.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at

<https://doi.org/10.1257/jep.33.1.209>

doi=10.1257/jep.33.1.209

Table 1

Examples of Tech Companies That Have Hired PhD Economists

| | | | |
|--------------|--------------|-----------------------|-------------|
| Alibaba | Forkcast | LinkedIn | Redfin |
| AirBnB | Glassdoor | Lyft | Ripple |
| Amazon | Google | Microsoft | Rover |
| AppNexus | Granular | Netflix | Trulia |
| CoreLogic | Groupon | Nuna | Uber |
| Coursera | Houzz | Oath | Upwork |
| Dstillery | Huawei | OpenAI | Vericred |
| Didichuxing | IBM | Pandora | Visa |
| Digonex | Indeed | Pinterest | Walmart |
| eBay | ING | PoliticalSheepdog.com | Wealthfront |
| ECONorthwest | Intel | Prattle | Yahoo! |
| Expedia | Kensho | Quantco | Yelp |
| Facebook | Lending Club | Quora | Zillow |

journals comparable to that of economists working in business schools or economics departments. These research centers, at their best, provide frontier insights, some of which will guide the future direction of the company.

Many tech companies now recruit directly through the American Economic Association's Job Openings for Economists (JOE) platform, which is where much of the recruiting for PhD economists begins. During the 2017–18 academic year, 21 tech companies were hiring through the JOE website. To put this into context, there are roughly half as many tech companies hiring through JOE as there are policy schools. Taking into account the fact that many of these companies had multiple positions, the number of positions available for economists in tech companies likely exceeded those at policy schools.

Moreover, Table 2 shows that the number of tech companies with job postings at JOE has generally risen in recent years. As technology platforms play an increasing role in the economy, topics relevant to them have become more important to the business school curriculum and to academic research in business schools. Business schools have seen increased demand for faculty specializing in online platforms and digitization, as well as in areas crucial to understanding data analysis, such as experimental methods and machine learning. For example, groups in business schools that historically focused on operations research or management of information systems have recently begun to focus more on economic problems such as marketplaces, pricing algorithms, and empirical studies of economic questions.

These shifts are partially driven by a growing need to prepare MBA students for a career in the technology sector. For example, Amazon was the largest employer of Harvard Business School's most recent graduating class of MBA students. Corresponding to the shifting career paths of MBA students, recent additions to the Harvard Business School curriculum in the past few years include courses on experimental methods, designing online marketplaces, digital marketing, technology strategy, and data science. Stanford's Graduate School of Business has seen similar

Table 2

The Number of Tech Companies, Policy Schools, Business Schools, and Economics Departments with Positions for PhD Economists

| | <i>Tech companies</i> | <i>Policy school departments</i> | <i>Business school departments</i> | <i>Economics departments</i> |
|------------------------------|---------------------------|--------------------------------------|--|----------------------------------|
| <i>August 2014—July 2015</i> | 15 | 43 | 247 | 366 |
| <i>August 2015—July 2016</i> | 15 | 43 | 264 | 349 |
| <i>August 2016—July 2017</i> | 23 | 36 | 289 | 322 |
| <i>August 2017—July 2018</i> | 21 | 50 | 326 | 374 |

Note: Authors using data from Job Openings for Economists (JOE).

growth. More broadly, there has been a rapid expansion in courses directly related to the technology industry. Content related to the digital economy has increasingly been added to more traditional courses (such as core strategy and marketing courses) as well.

Within industry, there is little precedent for private companies recruiting academic economists as well as new PhDs with strong research skills so heavily for full-time positions. Organizations like the RAND Corporation and Mathematica Policy Research recruit economists on a large scale, but focus mainly on research and policy evaluations. Consulting firms like Cornerstone and the Analysis Group also recruit large numbers of economists, but primarily to support and serve as expert witnesses in legal matters in areas such as antitrust and intellectual property litigation.

Considering the tech firms that hire into research labs, such as Microsoft Research, perhaps the closest historical analog would be Bell Labs, which was operating as a division of AT&T when it created an economics team in 1968. The team grew to include about 30 economists, including high-profile economists such as Elizabeth Bailey, Roy Radner, and Robert Willig. In 1970, it launched the *Bell Journal of Economics and Management Science*, which lives on as the highly regarded *RAND Journal of Economics*. The team was phased out in 1983, coinciding with the breakup of AT&T. Some of its economists were folded into other parts of the company, while others left for other industry or academic jobs—including at Columbia University, Harvard Business School, New York University, Princeton University, and the University of Pennsylvania.

Although some tech companies hire economists using a lab model, the majority of economists in tech companies work on managerially relevant problems with data from the company, and many are in business roles. For example, outside of Microsoft Research, Microsoft has a business-focused chief economist whose team actively recruits PhD economists to work on problems ranging from cloud computing to search advertising. Amazon assigns economists to specific business problems across divisions, ranging from the e-commerce platform to digital content to the experimentation platform used to evaluate changes and innovations. Uber has teams

of economists focused on understanding policy issues in addition to pricing and incentive design—some of these teams produce outward-facing research published in academic journals while others are completely inward-facing. More broadly, many economists at tech companies do a combination of external research and internal work, continuing to attend conferences and publish in leading economics journals; they often hire summer interns from top PhD programs or collaborate with academic economists on such projects. Because many of the problems faced by tech companies are on the frontier of academic research, close ties to academics and rigorous, original thinking are highly valued in the tech sector.

Indeed, the interaction between tech companies and economists has given rise to new intellectual questions and a new field within economics—the “economics of digitization.” The field has explored a wide range of questions. For example, how does the advent of artificial intelligence and the use of large-scale consumer datasets affect industry structure and market power? How should tech companies be regulated? How should data from the tech sector inform policy? How do aggregators, search engines, reputation systems, and social media affect the decisions we make and the news we read? How should online marketplaces be designed to ensure safe and efficient transactions? Online platforms have also created novel datasets and testing grounds that have been used to inform virtually every field of economics, from market design to industrial organization to labor economics to behavioral economics.

We have had the opportunity to spend our careers thus far with one foot in academia, studying and teaching about online platforms, and the other in practice, helping to shape them. Outside of our academic roles, we work closely with tech companies. Susan previously served as consulting chief economist at Microsoft and currently sits on the boards of Expedia, Lending Club, Rover, and Ripple. While working with Microsoft, she also helped build the economics group at Microsoft’s research arm in New England. Mike works with a variety of tech companies, and created an economic research initiative at Yelp. As academics, we have taught hundreds of students and executives who now work in tech companies. Doctoral students have become interested in tech companies as well—our own students have worked at companies ranging from Facebook, Microsoft, and Amazon to Wealthfront, Uber, and Airbnb.

The core skills that economists use in tech companies have been important to economic research for decades prior to the tech era. The field of market design has been combining novel theoretical insights, empirical work, and experiments to solve real-world problems since Bob Wilson’s pioneering work on auctions in the 1960s. Assessing causal relationships and understanding incentives have been central themes in applied microeconomics and industrial organization for decades. With the advent of new technologies, the expertise developed by PhD economists has found new and influential uses in the tech sector. Furthermore, the frontiers of economic research in these areas has been advanced as the tech sector has simultaneously introduced new economic problems, provided new ways to bring ideas from economic theory into practice, and provided opportunities for new types of statistical analysis.

With the rise of economists in tech companies, we're frequently contacted by tech companies for recommendations about whom to hire and what types of roles economists should take on. We are also asked how undergraduates and PhD students can prepare for such careers, as well as what these careers will be like. Faculty are often interested in how they can get involved with tech companies, and what types of problems they might work on there. In this paper, we describe 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.

What Tech-Relevant Skills Do PhD Economists Have?

To draw inspiration from Liam Neeson's line in the movie *Taken*, economists have "a very particular set of skills." Here, we focus on three broad skillsets that are part of the economics curriculum that allow economists to thrive in tech companies: the ability to assess and interpret empirical relationships and work with data; the ability to understand and design markets and incentives, taking into account the information environment and strategic interactions; and the ability to understand industry structure and equilibrium behavior by firms.

Assessing Empirical Relationships

Relative to other disciplines, economists have several strengths in thinking about data. First, economists are interested in understanding which relationships are *causal*—and which are not. Over the past 30 years, economics has developed a toolkit to identify causal relationships in real-world data. As the internet age has helped to usher in an era of unprecedented amounts of data, this has also contributed to the growing demand for economists.

For example, empirical applied microeconomics has developed tools for using "natural experiments" and for evaluating policies—tools such as instrumental variables, causal panel data models, and regression discontinuity (for a review of some of these, see Angrist and Pischke 2009). As we describe further in the next section, these tools are widely used in technology firms to answer questions about the effects of interventions such as price changes, the introduction of new products, changes to the user interface, and advertising effectiveness. Economists' attention to identifying causal effects, as well as to both the statistical and economic significance of findings, are important contributions to the practice of empirical analysis in tech firms. Industrial organization economists and market design economists have also developed methods for estimating the impact of counterfactual price changes or changes to market design. Perhaps surprisingly, these tools are less widely used in tech firms than the tools of empirical applied microeconomics, although there are notable exceptions.

Experiments are central to the decision-making process within the tech sector. Most large tech companies evaluate product changes through "A/B testing," or

randomized controlled trials, conducting thousands or tens of thousands of A/B tests per year. Experiments pose important managerial and technical questions, ranging from how to choose an appropriate sample, to how to design the intervention itself, to how to move from experimental results to a managerial decision.

With many experiments seeking to identify small effects over a massive number of users, changes to the methodology of A/B testing can be impactful. The science of experimental design has therefore become an important topic within tech companies, often pushing the research frontier. For example, Blake and Coey (2014) highlight challenges in running experiments in marketplaces, where equilibrium effects create interference between treatment and control groups in a paper motivated by challenges they faced at eBay and Facebook. Athey, Eckles, and Imbens (2018) examine issues that arise in evaluating experiments in a network setting, in a paper motivated by challenges they faced at Amazon and Facebook.

The widespread use of experiments in the tech sector has at times proved controversial, as when Facebook ran an experiment to test how users would react when Facebook varied whether users were shown more positive or negative posts in their newsfeeds (Kramer, Guillory, and Hancock 2014). Although the experiment ultimately found very small effects, it generated considerable public backlash against Facebook (Meyer 2014) and an expression of editorial concern from the journal that published the experiment (Verma 2014). In response to public pressure and broader concerns about the ethics of experimentation within companies, Facebook updated its internal procedure for deciding which experiments to run. Companies and policymakers are still exploring ways to establish best practices that allow for productive experimentation and uses of data, while protecting the privacy and safety of participants.

The widespread use of machine learning in tech firms has also created new challenges and opportunities. Initially, academic economists were slow to take up machine learning for reasons ranging from the lack of asymptotic results behind many approaches to machine learning to questions about whether prediction problems are important from an economics perspective. Thus, some economists came to tech firms unfamiliar with machine learning, requiring them to learn a new set of methods in order to communicate with the machine learning community. More recently, the interaction of economists with technology firms has contributed to an expansion of interest among economists in machine learning—focusing both on prediction problems and causal inference problems.

Motivated by the need to bring causal inference techniques to the large datasets of technology firms as well as the desire to make full use of these rich datasets, a recent literature has developed combining machine learning and causal inference (Athey, forthcoming), and this literature in turn has influenced the business practice of technology firms: for example, Hitsch and Misra (2018) apply Wager and Athey's (forthcoming) causal forest method in an application to targeted promotions, while the Athey and Imbens (2016) approach to recursive partitioning for causal effects has been applied in technology firms' A/B testing platforms. From a practical perspective, the intersection of machine learning and economics allows economists to understand what works, what doesn't, and why.

While experiments have played an important role within tech companies, they also have limitations. Economists have helped to bring a broader causal inference toolkit to supplement experiments within tech, using methods such as instrumental variables, causal panel data models, and regression discontinuity. This has allowed companies to obtain treatment effects in contexts where experiments might be difficult or costly to run.

In addition to their focus on causal relationships, economists are interested in understanding the tradeoffs involved in different outcome metrics. In many technology firms, decisions about product design, marketing, and even human resources are determined by empirical analysis (rather than subjective evaluation), and the choice of metrics will guide incentives throughout the companies. Economists have sought to understand the relationship between short-term metrics such as clicks on an advertisement (also called “surrogates,” as in Athey, Chetty, Imbens, and Kang 2016) which are easy to observe, and longer-term metrics (like revenue or the lifetime value of a customer), which are more difficult to observe, but better represent company goals.

For example, a large technology company made the following change in measurement for email marketing. The old measure, customer sales, was noisy. Consumers might take weeks before making a purchase. The new measure, opening the email, was immediately observable, and could be incorporated to adjust the content of the email very quickly. However the company found that within months, the number of sales *per email* declined precipitously, because the marketing emails evolved to maximize email opening rates without regard to final sales. For example, the successful emails (using the opening rate metric) had catchy subject lines and somewhat misleading promises. For economists, it is natural to think about a metric not only as a statistical measure, but also as implicitly governing worker incentives, and to suggest ways to provide incentives for long-term innovation as well as short-term metrics that better capture long-term effects. More broadly, economists are interested in the difference between short-term and long-term objectives, which can often lead to dramatically different conclusions in making product and market design choices, and in developing algorithms. Economists have focused on the link between experiments, algorithms, and managerial decisions.

Finally, the theoretical and empirical training of economists prepares them to think about both intended and unintended consequences of different decisions. For example, Airbnb originally made it very easy for landlords to reject guests after seeing their name and picture. While this extent of flexibility may have led to short-term user growth (the metric that Airbnb had been focusing on), an experiment run by Edelman, Luca, and Svirsky (2017) showed that it also led to widespread racial discrimination against African-Americans on the platform. Thus, Airbnb’s market design choices raised the possibility of reintroducing discrimination to a market that had worked hard to reduce it. Fisman and Luca (2016) proposed a series of market design choices that might reduce discrimination in online markets more generally—such as further automating transactions on platforms. As a result of this work, the company created a task force that weighed the different options, which

led to a full-time team of data scientists to explore discrimination on an ongoing basis. Regulators also became involved, prompting Airbnb to continue these efforts. Ultimately, Airbnb implemented a variety of changes that balanced their desire for short-term growth against the company's goal of reducing discrimination on the platform, objectives which were not always aligned.

Designing Markets and Incentives

The rise of economists in tech companies has coincided with the rise of market design, a field that was pioneered by Stanford economist Bob Wilson and extended into a variety of application areas by economists such as Paul Milgrom and Al Roth (who won the Nobel Prize for his pioneering work in this field). Market design has shifted economists away from using a primarily descriptive lens to a more prescriptive one, using the tools of economics to engineer better-functioning markets. These economists—and Roth in particular—have promoted the idea of the “economist as engineer,” whereby the economist becomes deeply involved in the implementation of economic ideas and tailors recommendations to the fine details of the problem. While market design research initially focused on offline marketplaces such as spectrum auctions, residency matching programs, and kidney exchange, economists have more recently taken the market design mindset into the tech sector. For example, the lens that Roth has long used in offline markets—exploring issues around market thickness, congestion, and safety of participants—has gained further prominence in online marketplaces, where design choices are front and center.

Applications of market design in tech firms range from Google, Yahoo!, and Microsoft's marketplaces for selling advertisements (Varian 2007; Edelman, Ostrovsky, and Schwarz 2007; Athey and Ellison 2011; Agarwal, Athey, and Yang 2009; Athey and Nekipelov 2013) to Uber's market for rides (Cohen, Hahn, Hall, Levitt, and Metcalfe 2016). Much of this literature has examined pricing and allocation mechanisms, as well as reputation systems. Other work has focused on search costs (Athey and Ellison 2011; Fradkin 2017; Cullen and Farronato, 2018). Multisided platforms are especially ripe for an economist's skills, since these are exactly the kinds of settings in which it is critical to think through strategic behavior, interactions, and equilibrium effects.

Bringing together their unique perspectives on assessing empirical relationships with their expertise in market design, economists offer particular value to technology firms by bringing together theory and data to predict not just the immediate effect of a decision, but how a decision affects equilibrium behavior in a market.

Analyzing Equilibrium Market Structure

Tech companies think a lot about which markets to enter, taking into account the current and potential competitive landscape. For example, questions about market structure have arisen in the battle between Uber and Lyft, and helped to shape expansion and acquisition strategies. Economic theory, including the theory of platforms and market design, speaks to the forces that might lead a market to

be highly competitive, as well as the forces that make monopoly more likely. This is helpful for platforms deciding a strategy about which markets to enter, and also for policymakers and regulators. Currently, the question of market power is hotly debated in the technology industry, and economists can help by putting structure on the debate, even if they cannot perfectly predict the future.

Applications of Economics in Technology Firms

Economists now work on a variety of issues pertaining to tech companies. In this section, we highlight several exemplars of economics in tech companies: designing advertising auctions, estimating the returns to advertising, designing review and reputation systems, and studying the effects of reviews on firms.

Design of Online Advertising

Advertising has changed dramatically with the advent of online technology, and with the involvement of economists. This involvement has been concentrated in two areas: the design of advertising auctions and estimating the returns to advertising.

The involvement of economists in online advertising auctions dates back to the late 1990s, when Simon Wilkie, an economics professor at Cal Tech, started advising GoTo, a company that later became Overture and eventually powered Yahoo!'s search advertising auctions. In 2002, Hal Varian received a call from Eric Schmidt, the chairman of a young company called Google. Schmidt was intrigued by *Information Rules*, a book Varian had coauthored with Carl Shapiro, his fellow economist and colleague at the University of California, Berkeley. After speaking with Schmidt, Varian became a consultant for Google, and ultimately, the company's chief economist, the first academic microeconomist to become chief economist of a major technology firm. Preston McAfee, another market design economist, joined Yahoo! Research from Cal Tech a few years later, while Susan became consulting chief economist at Microsoft while on leave from Harvard in 2008. Susan and Preston McAfee also initially focused on market design and strategy questions surrounding online advertising.

To understand some of the issues involved with search advertising, consider the way it works. Search engines, ranging from general engines like Google and Bing to more specialized search engines like Yelp, generally sell advertising through auctions for specific terms. Bids are expressed in terms of a willingness to pay per outcome, such as a click, and advertisers with higher bids are rewarded with more favorable ad placement. Thus, firms must make choices about auction formats and parameters.

One complexity arises because in a traditional second-price auction with a single winner, the winner pays the bid of the second-highest bidder, which in turn means that the best strategy is to bid one's true value (and not to worry about being an outlying high bidder). However, in ad sales, the result is a ranking of bidders, not a single winner. The auction used by Google, Microsoft, and Yahoo! is a generalized

second-price auction, where each advertiser pays the price bid by the next lowest bidder. Work by Edelman, Ostrovsky, and Schwarz (of Yahoo! Research) (2007) shows that the generalized second-price auctions do not have the same properties as a second-price auction with a single winner (for instance there is not a dominant strategy equilibrium), but they remain useful in search engine advertising applications.

Athey and Ellison (2011) incorporate rational consumer search into the market design of auctions, motivating the use of reserve price not only as an instrument for raising revenue, but also as a tool for managing advertising quality and thus increasing users' incentive to search. One of us, Susan, used this as a framework for advising Microsoft to improve the ad quality on Microsoft's search engine. Later she took the theoretical models to the data and built an econometric model (Athey and Nekipelov 2012) that could be used to infer advertiser valuations and profits from their bidding behavior. This type of model can be used to understand how changes in algorithms affect advertiser well-being and thus forecast the future engagement of advertisers on the platform.

At Yahoo!, Ostrovsky and Schwarz (2016) observed that the reserve prices the company was setting were lower than what auction theory predicted would be revenue-maximizing for the seller. The pair assigned search keywords to a treatment and a control group. Keywords in the treatment group received a theoretically optimal reserve price calculated by the authors, while keywords in the control group used a default reserve price of \$0.10 per click. The treatment group increased ad revenue by several percentage points, leading Yahoo! to change its reserve price policies for all of its search advertising—and making the company millions of dollars in additional revenue.

Tech firms have also hired economists to solve challenges relating to the choice of outcome of advertising, such as pay-per-click versus alternatives. Agarwal, Athey, and Yang (2009) explore the benefits and drawbacks of pay-per-click compared to pay-per-action, in which advertisers only pay each time an individual performs an action after clicking the ad link—such as buying a product. Pay-per-action also allows advertisers to better express their benefits from different ad placements; some ad placements may attract consumers who are likely to purchase expensive items, but other ad placements may attract consumers who purchase cheap items, or do not purchase at all. Pay-per-action allows the advertising platform to optimize on behalf of the advertiser, obtaining more placements in scenarios where more profitable consumer behavior is likely. However, if the pay-per-action system allows bidders to bid on several different types of actions, the bidder may have opportunities to game the system, potentially making the revenue to the search engine arbitrarily low.

Finally, although not much academic work has analyzed Facebook's online advertising auctions, Facebook's early decision to adopt a Vickrey auction to sell its advertising space was heavily influenced by the training of a Facebook employee, John Hegeman, in the graduate program at the economics department at Stanford, which has considerable expertise in auctions (Amit, Grefl, and Hegeman 2013).

The Role of Ranking and Incentives in Marketplaces

Equilibrium effects can be especially challenging to understand in the platforms and marketplaces that are common in the tech industry. For example, a change to the user interface at eBay that made it easier for consumers to find the products they want, and thus to do price comparisons, affected consumer choice behavior (Dinerstein, Einav, Levin, and Sundaresan 2018), but that in turn can affect the prices charged by sellers. Over the long term, changes in pricing behavior by sellers affect consumers' desire to shop on eBay at all, which in turn influences seller behavior. Similar issues arise in many marketplaces. In general, the way marketplaces and intermediaries rank the offers from sellers or service-providers can be thought of as an incentive system. Marketplaces like Airbnb incentivize owners to maintain their calendars accurately and accept booking requests from travelers by prioritizing owners who behave as desired, and demoting those that do not. Economists are well positioned to analyze issues that arise in ranking offers from sellers, not just on short-term user behavior, but also the equilibrium impact on the marketplace as a whole.

Estimating the Returns to Advertising

Estimating the returns to advertising has traditionally been difficult. Older media, such as print and television, do not allow for showing different advertisements or tracking behavior at the individual consumer level, which makes designing randomized experiments difficult. Nonrandomized observational studies are biased due to selection issues. Thus, most traditional studies of advertising were plagued by poor identification strategies, limited data on outcomes, and small sample sizes.

The digital age has allowed a better understanding of the returns to advertising. Platforms such as Facebook, Google, and Microsoft collect vast amounts of data on user behavior, and regularly run experiments to test the effectiveness of their online advertising systems—allowing them to make progress on understanding the conditions under which advertising is most effective. Economists at such firms can thus draw on existing theories of market design, generate new ideas, and rapidly test and evaluate those ideas.

Economists at companies that advertise online have also made significant progress in understanding the effectiveness of advertising. For example, while working at eBay Research Labs, Blake, Nosko, and Tadelis (2015) conducted field experiments that allowed them to understand the impact of eBay's advertising campaigns on Google and Bing. They found that search engine marketing—purchasing ads to be displayed on search engines when certain search terms are entered—was only effective when ads are viewed by new or infrequent eBay customers and when the search terms already contain the firm being searched for. Since frequent customers drive most of their sales, the overall returns were negative, a significant result given that eBay's yearly US search engine marketing budget was over \$50 million at the time of the experiment.

In other contexts, advertising appears to be a positive investment. Johnson, Lewis, and Reiley (2016) report a 3.6 percent increase in sales among consumers

shown advertisements for a large retailer on Yahoo!, with a point estimate, though not statistically significant, of positive returns. Their experiment used a sample size in the millions, a control group shown an irrelevant ad (in addition to a group shown no ads), and a large set of individual covariates. Dai, Kim, and Luca (2016) collaborated with Yelp to display ads randomly for a set of previously non-advertising restaurants—a design that allowed them to include many small businesses rather than a small number of well-known businesses. Restaurants for which ads were shown had 25 percent more page views and 5 percent more reviews (which can be viewed as a proxy for actual visits to the restaurant)—and a back-of-the-envelope calculation suggests a positive return on investment.

Economists have also designed long-term experiments that examine the impact of ads on the propensity of users to buy or use the advertised product; Huang, Reilly, and Riabov (2018) study Pandora Internet Radio consumers over a 21-month period, estimating a fairly linear negative relationship between the quantity of ads shown to each consumer while listening to the internet radio and usage of the service, and further showing that increasing the ad load increased purchases of paid, ad-free subscriptions.

But it remains challenging to measure the returns to advertising. Lewis and Rao (2015), two economists formerly at Yahoo!, discuss the challenges in a meta-analysis spanning 25 online advertising field experiments. They argue that even studies of returns to advertising that can use online data are still held back by the signal-to-noise ratio in individual sales data, where standard deviations are often an order of magnitude higher than means. Even studies with hundreds of thousands of users often produce confidence intervals too wide even to distinguish highly profitable ads from wholly ineffective ones.

Designing Review and Reputation Systems

Online reviews and reputation systems have become increasingly prevalent in the past decade. Platforms like Yelp and TripAdvisor contain hundreds of millions of reviews for businesses ranging from plumbers to hotels. Uber, Airbnb, and other online marketplaces also rely heavily on reputation systems to facilitate trust between strangers, and traditional retailers ranging from Home Depot to Gap have developed review systems of their own.

Economists have been involved in the design of reputation systems—focusing on understanding the systematic biases that can occur in review ecosystems and the design choices that might mitigate these biases. One bias that has been documented in review systems in online marketplaces arises from reciprocal reviewing, in marketplaces where buyers and sellers review each other. While reciprocal reviewing can be a valuable way to build trust on both sides of the market, it has the potential to create incentives for upward-biased reporting. When Airbnb allowed the reviews of renters to be posted before those of the hosts, guests might have been hesitant to leave bad reviews out of concern that hosts would reciprocate. Bolton, Greiner, and Ockenfels (2013) propose a fix to this dilemma in the context of eBay, which offered reciprocal reviewing where both buyer and seller

reviews were immediately posted. The solution eBay (and Airbnb) explored is to postpone displaying reviews until both sides have left a review, or until a certain amount of time has expired. Under this system, however, buyers may still be reluctant to provide negative feedback if they suspect that it would discourage future sellers from transacting with them. Therefore, eBay added an anonymous, one-way review component called a “detailed seller rating,” where buyers assign sellers several numerical scores and the results are only viewable in aggregate form. Fradkin, Grewal, and Holtz (2018) study this issue using a randomized experiment at Airbnb (working within the company), and find results consistent with the hypothesis that reducing the possibility of retribution increases the informativeness of reviews.

A second bias can arise because reviews in online marketplaces are voluntary and so may suffer from selection bias. In particular, users may be more likely to leave a review after an especially positive or negative experience. For example, a group from eBay’s team of economists found evidence that eBay users were more likely to leave a review after a positive experience, relative to a negative one (Masterov, Meyer, and Tadelis 2015). Review platforms have a variety of tools to tackle the selection problem, such as sending emails to encourage consumers to leave reviews and even paying reviewers. Alternatively, platforms can incorporate information about buyer and seller review frequency into reputation scores—for example, penalizing sellers who receive low rates of feedback. Upon the recommendation of an in-house economist, a large online labor market developed a system that allowed for both private and public feedback, finding that private feedback was less inflated than public-facing reviews.

A third bias in online reviews occurs when businesses, or individuals hired by businesses, surreptitiously leave reviews about themselves or their competitors. Luca and Zervas (2016) explore the role of economic incentives in a business’s decision to commit review fraud, finding that independent restaurants and restaurants with a declining reputation are more likely to commit review fraud, and restaurants with high competition are more often targeted with fake negative reviews. One mechanism for reducing fraudulent reviews is to verify whether a transaction has occurred before allowing a review, as is policy on Airbnb, for example; other sites, such as Amazon, label reviews that come from a verified purchase. While this precaution may reduce fake reviews, it may also prevent legitimate reviews on some platforms by increasing the barriers to contributing content. Mayzlin, Dover, and Chevalier (2014) find evidence of promotional reviews in the context of TripAdvisor (which does not verify that reviewers have stayed at a property) and Expedia (which does). They find that relative to chains, independent hotels tend to have more five-star reviews on TripAdvisor (relative to Expedia). Moreover, the competitors of independent hotels tend to have more one-star reviews on TripAdvisor relative to Expedia.

In addition to creating incentives for people to leave high-quality reviews, platforms face a problem of how to aggregate reviews once the reviews are in place (Dai, Jin, Lee, and Luca 2018). In practice, review platforms such as Yelp and TripAdvisor

use algorithms to identify and remove content that is thought to be fake or of low quality. Platforms can also adjust and weight ratings to account for the informational content of each review, increasing the overall informational content of average ratings being presented to users. In practice, platforms also have to consider the incentive effects that different approaches to filtering and aggregating content might have.

Another perspective on reviews that is natural from an economist's training is to consider the cost of a user's time in writing a review as balanced against the value of information from a review. For example, Uber makes a decision about whether to require all riders to leave a review, or whether to request reviews only in some situations. It may not be worthwhile to request a review from every user who interacts with a highly experienced and well-rated seller on the marketplace. On the other hand, it is important to continue to collect some reviews to provide continued incentives for the seller to provide quality. In addition, there may be aspects of the user experience that can be directly measured by a marketplace (for example, time it took for the seller to ship, whether an Uber rider exceeded the speed limit, or how much a rider tips the driver). In such cases, it may be more efficient to ask the buyer to review aspects of the service that are more difficult to observe or infer directly.

The Effects of Reviews

The effects of online reviews on demand for products can be hard to identify. For example, hotels with higher TripAdvisor ratings may have higher demand either because ratings drive demand or simply because better hotels have higher ratings. However, economists have used a variety of methods to identify the causal impact of online reviews.

As one example, consider a book that is sold both on Amazon and on the Barnes & Noble website. The book would almost certainly have different ratings on the two platforms. Moreover, if an Amazon user left a review, the rating would change on Amazon, but not on Barnes & Noble, leading to variation in ratings across platforms and over time. Arguing that the exact timing of incoming reviews is plausibly exogenous, Chevalier and Mayzlin (2006) use this variation to estimate the impact of reviews on online book purchases. Specifically, they look for increases in sales on Amazon (relative to Barnes & Noble) after a review was left on Amazon (but not on Barnes & Noble)—implementing a difference-in-differences strategy. Using a regression discontinuity approach, Luca (2016) finds that higher ratings lead to higher sales for independent restaurants, but finds no evidence of this for chains. Anderson and Magruder (2012) find similar effects of Yelp ratings on restaurant reservations. Ghose, Ipeiritos, and Li (2012) uses a similar approach to understand the impact of TripAdvisor reviews. Beyond the average rating, other aspects of reviews are potentially important. For example, Sun (2012) explores the impact of the variance of product reviews, and highlights that if the variation in reviews of a product is driven by heterogeneous preferences, then, holding fixed the average review of a seller, it may be better to match some customers with sellers who have more variable reviews—as the variation may reflect the fact that the product is a good match for some customers but not for others.

Consumer reviews also have important implications for market structure and consumer welfare. Clemons, Gao, and Hitt (2006) argue that information provided in reviews can help to grow demand for products that are more differentiated by increasing the quality of the match, and find evidence generally consistent with this argument when looking at reviews for beer and growth in demand. Bar-Isaac, Caruana, and Cuñat (2012) theoretically show that introducing new information into a market can lead to a higher degree of product differentiation in markets. This finding suggests that the existence of online reviews may lead to a greater variety of products and services. Lewis and Zervas (2018) estimate the welfare effects of TripAdvisor reviews, focusing on the reduced search costs in markets with more review content.

Acquisitions, Exclusive Deals, and Strategy

The first question Susan was asked at Microsoft was whether internet search with search advertising was an industry that could sustain two or three players, or whether it was destined to be a monopoly. Her analysis of scale economies and indirect network effects in search suggested that sufficient scale was necessary for a second search engine to be viable; this analysis was used to value Microsoft's acquisition of Yahoo!'s search business, as well as other large business deals involving search. Later, the question arose of whether the smartphone market could sustain three platforms, something that has proved difficult to achieve. Questions about vertical integration also arise in these markets; for example, Google acquired the ITA travel search engine in 2010. Prior to that, ITA was providing the technology powering the travel search results for Microsoft's competing search engine, setting the stage for Google to increase dramatically its share of travel searches. This acquisition was closely reviewed by the US Department of Justice and was eventually approved with certain conditions (Miller 2011). Later, the European Commission imposed large fines on Google for biasing search results in favor of its own vertically integrated specialized search services (Scott 2017), and later for tying its search engine and mapping services to the applications store for Android (Satarino and Nicas 2018). Banks around the world have complained that Apple gives the Apple Wallet exclusive access to the NFC radio, a crucial component of mobile payments, in the iPhone. Apple then takes a fee for every credit card transaction that takes place on the Apple Wallet, a fee that is large (up to 0.15 percent) relative to the profits of the credit card networks (Zhu, Athey, and Lane 2018). Banks faced difficult strategic questions about whether to enable Apple Wallet in light of these fees as well as the control they would give up to Apple. Tech economists have been involved in analyzing all of these issues from both a business and regulatory perspective.

Economic theory and empirical approaches can also be critical in analyzing exclusive deals in the tech industry. For example, when gaming platforms such as Microsoft's Xbox and Sony's Playstation release new generations, they typically sign exclusive deals for games. Economic theory and empirical methods (for example, Lee 2013) can be used to value these exclusive deals, incorporating the direct impact

of those games on the sales of consoles at the time of launch, but also the indirect effect of those additional consoles on the subsequent incentives of game developers to develop for a platform, which in turn affects consumers, and so on.

Positions for Economists at Tech Companies

Economists have had mixed reception in tech companies. While some companies like Amazon have been quick to bring economists into the highest levels of decision-making, others have been slower, with economists sitting within data science teams or policy teams with less influence over the direction of the firms. In practice, economists within technology companies take on a number of roles ranging from Chief Economist to Product Manager. Economists often work within inward-facing groups at companies, including forecasting and planning, pricing, testing, and data science teams as well as outward-facing groups including policy, public relations, and marketing teams. We outline some examples of these roles.

Data science/analytics is one of the fastest-growing job areas as tech companies become more data-driven. Economists use observational and experimental data to answer business questions, such as whether to introduce new products, how to understand the effectiveness of large initiatives, and how to evaluate the impact of competitors. Because this work directly informs the decisions of many other departments, some firms have embedded data scientists in product teams while others have centralized data science teams. For example, Amazon currently embeds data scientists within product teams, while Yelp has a centralized data science team. Economists often help to manage teams of data scientists as well—for example at Coursera, or for a period of time at HomeAway.

Tech companies are increasingly using *experimentation or A/B testing* to answer product or platform design questions, such as the launch of a new product or advertising campaign. Economists can help to manage the design, process, and analytics around randomized controlled experiments. Some firms have embedded A/B testing specialists within their functional teams (for example, in marketing teams) while others have a separate team to manage a larger testing platform. For example, Uber and Facebook have economists involved in managing experimentation platforms and process in a context with strong network effects and many experiments. Other economists have developed and applied techniques for estimating heterogeneous treatment effects in A/B testing platforms (for example, Athey and Imbens 2016; Wager and Athey forthcoming).

Some tech firms have embedded experimentation or data scientists into their *advertising/marketing analytics*. These teams typically evaluate the effectiveness of advertising, design experiments around advertising, optimize advertising spending, and predict the success of advertising campaigns. For example, Netflix has a team working on these issues.

Economists working as *product managers* can also design experiments and surveys that answer questions that guide product designs and other strategic decisions,

including ranking algorithms in search platforms or presentation of information in stores. These tasks often involve drawing causal inferences from observational data—for example, using difference-in-differences methods to evaluate the impact of a new product or feature.

In *regulation/litigation* settings, the role of an economist includes writing policy white papers that translate theory and empirical work for a legal or policy audience, contributing knowledge of specific subject areas such as telecommunications policy, intellectual property, and antitrust from an economic perspective. Chief economists often spend a share of time on these issues as well. Airbnb has economists trying to understand housing markets and policy. Uber has economists investigating the impact of Uber on the taxi industry and quality of rides. Google (and previously, Yahoo! and Microsoft) has had economists studying antitrust issues related to Google's dominant position in the search industry.

Tech companies also have economists in a *public policy role*, helping to partner with policymakers, often through data-sharing and analysis. For example, Yelp partnered with the City of Boston to develop an algorithm that allowed the city to help target inspections for restaurant health code violations (Glaeser, Hillis, Kominers, and Luca 2016). Yelp data has been used to forecast government statistics (Glaeser, Kim, and Luca 2017), to understand how neighborhoods change during gentrification (Glaeser, Kim, and Luca 2018), and to estimate the impact of the minimum wage on restaurant exit and prices (Luca and Luca 2018). Yelp also partnered with cities (and a third party data provider) to display hygiene violations online, providing a modern, digital implementation of the hygiene disclosure policies analyzed by Jin and Leslie (2003), where regulation forced restaurants to prominently display their hygiene ratings in their stores. This initiative helped to steer customers away from restaurants with the most violations of health code policy (Dai and Luca 2018). Zillow, the online real estate company, creates reports of local housing markets. Search data from Yahoo! and Google has been used to help forecast economic activity (Goel, Hofman, Lahaie, Pennock, and Watts 2010; Choi and Varian 2012; Wu and Brynjolffson 2015). LinkedIn is exploring the ways in which its data can help to shed light on labor markets. Uber's public policy team examines issues such as the impact of driving for Uber on driver welfare (Chen, Chevalier, Rossi, and Oehlsen 2017), the impact of Uber on labor markets and local economies (Hall, Horton, and Knoepfle 2017), and the role of gender in labor markets (Cook, Diamond, Hall, List, and Oyer 2018).

Several leading technology companies, including Zillow and Houzz, employ economists to do research designed for *public and media relations*, intended to inform potential customers and to create awareness for the company. For example, a primary mechanism for Zillow to attract consumers in its early years was that its chief economist created analyses of real estate markets to be covered by local and national news media. As another example, Houzz employs PhD economists who analyze and publish trends and data relevant to home remodeling.

Members of the *chief economist team* conduct and oversee many of the roles outlined above and also may make strategic decisions for the company. These decisions might include acquisitions and partnerships (one of us, Susan, worked on

strategy and empirical analysis for Microsoft's investment in Facebook, the acquisition of Yahoo!'s search business, and the company's strategy for cloud computing), as well as pricing and market entry.

Depending on the size of the company, economists have also gone into forecasting and planning teams (using time-series econometrics and modeling), pricing teams (using market design and supply-and-demand modeling), and academic relations (recruiting academics to fill the economic roles and build academic awareness surrounding policy and public relations issues).

Discussion

While we have focused mainly on economists working directly in tech firms, the rise of tech companies and emergence of the economics of digitization has important implications for academia as well. The shifting field leads not only to new research questions, but also to new academic positions, opportunities for collaborations, and potential career shifts. In this section, we address these opportunities.

Partnerships with Academics

While a growing number of economists now work within tech companies, collaborations with academics remain central to the strategy of tech firms and to the diffusion of economics within companies. For example, Airbnb, Amazon, eBay, Facebook, Indeed, LinkedIn, Microsoft, Rover, TaskRabbit, Uber, Upwork, Yelp, and Zillow have all collaborated with academic economists. These collaborations have several advantages for companies.

First, academics often have deep expertise in focused areas, including the key areas highlighted in this paper and many others; for example, a behavioral economist might shed light on the role of habit formation in user behavior. A market design economist might have unique insight into mechanisms driving market thickness. An econometrician might offer new ways to run experiments in a market with complicated network effects. Academics are also well-positioned to draw on insight from different contexts, since their work is less concentrated on a single platform.

Second, economists working full time within companies are often under pressure to deal with immediate issues (such as whether to change prices in a given quarter, or whether a specific advertising campaign was productive). Academics are insulated from these pressures, and so can explore longer-term strategic issues such as whether a company is even tracking the right metrics, or whether it makes sense to shift product composition.

Third, the hiring of economists by tech companies has brought forth a related risk—little research is being conducted internally on the shortcomings of tech companies and the negative implications of their models. For example, Airbnb did not examine racial discrimination on the platforms until academics documented it in academic research, thus bringing it to the attention of policymakers, Airbnb users, and ultimately, Airbnb managers. Working with academics and allowing a

broad degree of autonomy can help to get more credible and objective assessments of issues with which the companies are dealing.

At the same time, challenges do arise with academic partnerships. For example, academics often sign agreements with firms guaranteeing the ability to publish their results regardless of the result. In principle, this helps to reduce publication bias. However, firms may choose not to sign agreements around research topics where they are concerned about what the answers might be, potentially creating a bias towards papers favorable towards firms and creating an incomplete snapshot of an issue. This issue is not new, since economists have obtained data from firms and government agencies at their discretion for many years. However, as collaborations become more standard, this issue becomes more important.

Academic Jobs for Digitization Economists

The number of academic positions for digitization economists is growing. While some of these are in economics departments, digitization economists now also teach in business schools in strategy, marketing, information systems, entrepreneurship, and other departments. Doctoral students with interests in these areas should be aware that while recruiting for some of these positions takes place through the American Economic Association, recruiting for other positions, such as in marketing, operations, and information systems, often takes place on other timelines and outside of AEA mechanisms.

Tech companies have also created strong demand for undergraduate economics majors, who take roles ranging from product management to policy. Leading universities including Dartmouth, Harvard, MIT, Princeton, Stanford, and Yale teach about online platforms in their introductory microeconomics courses, or have created entire courses related to the “economics of digitization” (including courses on e-commerce, online platforms, and related areas). MIT’s economics and computer science departments have partnered to create a new major in computer science, economics, and data science. Harvard, MIT, and other universities have developed data science initiatives, drawing in computer scientists, economists, and other social scientists. We see opportunities to expand these course offerings, and to combine them with additional course material for students looking for a career at tech companies. Courses about marketplaces and platforms, taught from an economics perspective, have also proliferated among business schools, such as Boston University, Harvard, New York University, and Stanford.

While PhD economists are well suited to tech careers in many ways, we also see areas for the field to improve the preparation of PhD economists for working with or in tech companies. First, with the importance of prediction, targeting, and precise estimates in tech companies, machine learning plays an important role in tech companies. While the field of economics has long been a leader in causal inference, the field is still in the process of incorporating machine learning into its standard toolkit. Second, economists have historically received less training, relative to computer scientists, at coding and at optimizing code to run statistical algorithms at large scale. Investing in these skills (and incorporating them into the

PhD curriculum) can help to prepare economists to work in this area. At the same time, it remains important that economists have a strong conceptual understanding of economic issues like incentives and equilibrium effects, as well as strong empirical skills in the areas such as causal inference that we have described in this paper.

Shifts Between Academia and Practice

Economists in this area have growing opportunities to shift between academia and practice. Microsoft, Google, Yahoo!, Facebook, Amazon, eBay, Yelp, Uber, and other companies have all hosted faculty during sabbaticals. Tenured faculty members have left academia for positions at Amazon, Google, and elsewhere. Practitioners have also transitioned into academia—for example, leaving Facebook and Microsoft for MIT and Stanford. We believe this is the beginning of a larger movement in which a greater share of academic economists spend time in practice, acquiring a deeper understanding of what issues are most important for efficiency and profitability in technology firms, as well as getting exposure to unsolved business problems that may highlight fruitful academic research questions. As more PhD economists accept positions at tech companies, clearer paths for spending time (or re-entering) academia will likely appear, for those who are interested in this option. Firms that allow their economists to continue to publish will likely have an advantage in recruiting and retaining economists who want to retain flexibility in their career paths.

■ *We thank Duncan Gilchrist and Guido Imbens for valuable feedback. Stephanie Chan and Byron Perpetua provided excellent research assistance.*

References

- Agarwal, Nikhil, Susan Athey, and David Yang. 2009. "Skewed Bidding in Pay-per-Action Auctions for Online Advertising." *American Economic Review* 99(2): 441–47.
- Amit, Alon, Yaron Grefl, and John Hegeman. 2013. "Advertisement Selection and Pricing Using Discounts based on Placement." U.S. Patent 2013/0325585 A1 file June 4 2012, and published December 5, 2013. <https://patents.google.com/patent/US20130325585A1/en>.
- Anderson, Michael, and Jeremy Magruder. 2012. "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database." *Economic Journal* 122(563): 957–89.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton University Press.
- Athey, Susan. Forthcoming. "The Impact of Machine Learning on Economics." Forthcoming in *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay K. Agrawal, Joshua Gans, and Avi Goldfarb. University of Chicago Press.
- Athey, Susan, Raj Chetty, Guido Imbens, and Hyunseung Kang. 2016. "Estimating Treatment Effects Using Multiple Surrogates: The Role of the Surrogate Score and the Surrogate Index." Unpublished paper.
- Athey, Susan, Dean Eckles, and Guido Imbens. 2018. "Exact p -values for Network Interference."

Journal of the American Statistical Association 113(521): 230–40.

Athey, Susan, and Glenn Ellison. 2011. “Position Auctions with Consumer Search.” *Quarterly Journal of Economics* 126(3): 1213–70.

Athey, Susan, and Guido Imbens. 2016. “Recursive Partitioning for Heterogeneous Causal Effects.” *PNAS* 113(27): 7353–60.

Athey, Susan, and Denis Nekipelov. 2012. “A Structural Model of Sponsored Search Advertising.” https://people.stanford.edu/athey/sites/default/files/structural_sponsored_search.pdf.

Bar-Isaac, Heski, Guillermo Caruana, and Vicente Cuñat. 2012. “Search, Design, and Market Structure.” *American Economic Review* 102(2): 1140–60.

Blake, Thomas, and Dominic Coey. 2014. “Why Marketplace Experimentation is Harder than It Seems: The Role of Test-Control Interference.” In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, 567–582.

Blake, Thomas, Chris Nosko, and Steven Tadelis. 2015. “Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment.” *Econometrica* 83(1): 155–74.

Bolton, Gary, Ben Greiner, and Axel Ockenfels. 2013. “Engineering Trust: Reciprocity in the Production of Reputation Information.” *Management Science* 59(2): 265–85.

Chen, M. Keith, Judith A. Chevalier, Peter E. Rossi, and Emily Oehlsen. 2017. “The Value of Flexible Work: Evidence from Uber Drivers.” NBER Working Paper 23296.

Chevalier, Judith A., and Dina Mayzlin. 2006. “The Effect of Word of Mouth on Sales: Online Book Reviews.” *Journal of Marketing Research* 43(3): 345–54.

Choi, Hyunyoung, and Hal Varian. 2012. “Predicting the Present with Google Trends.” *Economic Record* 88(1): 2–9.

Clemons, Eric K., Guodong Gao, and Lorin Hitt. 2006. “When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry.” *Journal of Management Information Systems* 23(2): 149–71.

Cohen, Peter, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. 2016. “Using Big Data to Estimate Consumer Surplus: The Case of Uber.” NBER Working Paper 22627.

Cook, Cody, Rebecca Diamond, Jonathan Hall, John A. List, and Paul Oyer. 2018. “The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers.” NBER Working Paper 24732.

Cullen, Zoë, and Chiara Farronato. 2018. “Outsourcing Tasks Online: Matching Supply and Demand on Peer-to-Peer Internet Platforms.”

<https://www.hbs.edu/faculty/Pages/item.aspx?num=50051>.

Dai, Weijia, Ginger Jin, Jungmin Lee, and Michael Luca. 2018. “Aggregation of Consumer Ratings: An Application to Yelp.com.” *Quantitative Marketing and Economics* 16(3): 289–339.

Dai, Weijia (Daisy), Hyunjin Kim, and Michael Luca. 2016. “Effectiveness of Paid Search Advertising: Experimental Evidence.” Harvard Business School Working Paper 17-025.

Dai, Weijia (Daisy), and Michael Luca. 2018. “Digitizing Disclosure: The Case of Restaurant Hygiene Scores.” Unpublished paper.

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan. 2018. “Consumer Price Search and Platform Design in Internet Commerce.” *American Economic Review* 108(7): 1820–59.

Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. “Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment.” *American Economic Journal: Applied Economics* 9(2): 1–22.

Edelman, Benjamin, Michael Ostrovsky, and Michael Schwarz. 2007. “Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords.” *American Economic Review* 97(1): 242–59.

Fisman, Raymond, and Michael Luca. 2016. “Fixing Discrimination in Online Marketplaces.” *Harvard Business Review* 94(12): 88–95.

Fradkin, Andrey. 2018. “Search, Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb.” Available at SSRN: <https://ssrn.com/abstract=2939084>.

Fradkin, Andrey, Elena Grewal, and David Holtz. 2017. “The Determinants of Online Review Informativeness: Evidence from Field Experiments on Airbnb.” Available at SSRN: <https://ssrn.com/abstract=2939064>.

Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li. 2012. “Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content.” *Marketing Science* 31(3): 493–520.

Glaeser, Edward, Andrew Hillis, Scott Duke Kominers, and Michael Luca. 2016. “Crowdsourcing City Government: Using Tournaments to Improve Inspection Accuracy.” *American Economic Review* 106(5): 114–18.

Glaeser, Edward L., Hyunjin Kim, and Michael Luca. 2017. “Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity.” NBER Working Paper 24010.

Glaeser, Edward L., Hyunjin Kim, and Michael Luca. 2018. “Nowcasting Gentrification: Using Yelp Data to Quantify Gentrification.” *American*

Economic Association: Papers and Proceedings 108: 77–82.

Goel, Sharad, Jake M. Hofman, Sébastien Lahaie, David M. Pennock, and Duncan J. Watts. 2010. “Predicting Consumer Behavior with Web Search.” *PNAS* 107(41): 17486–90.

Hall, Jonathan V., John J. Horton, and Daniel T. Knoepfle. 2018. “Pricing Efficiently in Designed Markets: Evidence from Ride-Sharing.” http://john-joseph-horton.com/papers/uber_price.pdf.

Hitsch, Günter J., and Sanjog Misra. 2018. “Heterogeneous Treatment Effects and Optimal Targeting Policy Evaluation. Available at SSRN: <https://ssrn.com/abstract=3111957>.

Hu, Nan, Jie Zhang, and Paul A. Pavlou. 2009. “Overcoming the J-shaped Distribution of Product Reviews.” *Communications of the ACM* 52(10): 144–47.

Huang, Jason, David Reilly, and Nikolai Riabov. 2018. “Measuring Consumer Sensitivity to Audio Advertising: A Field Experiment on Pandora Internet Radio.” Available at SSRN: <https://ssrn.com/abstract=3166676>.

Jin, Ginger, and Phillip Leslie. 2003. “The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards.” *Quarterly Journal of Economics* 118(2): 409–51.

Johnson, Garrett A., Randall A. Lewis, and Elmar I. Nubbemeyer. 2017. “Ghost Ads: Improving the Economics of Measuring Online Ad Effectiveness.” *Journal of Marketing Research* 54(6): 867–84.

Johnson, Garrett A., Randall A. Lewis, and David H. Reiley. 2016. “When Less is More: Data and Power in Advertising Experiments.” *Marketing Science* 36(1): 43–53.

Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. “Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks.” *PNAS* 111(24): 8788–90.

Lee, Robin S. 2013. “Vertical Integration and Exclusivity in Platform and Two-Sided Markets.” *American Economic Review* 103(7): 2960–3000.

Lewis, Gregory, and Giorgos Zervas. 2018. “Welfare Impact of Consumer Reviews: A Case Study of the Hotel Industry.” Unpublished paper.

Lewis, Randall A., and Justin M. Rao. 2015. “The Unfavorable Economics of Measuring the Returns to Advertising.” *Quarterly Journal of Economics* 130(4): 1941–73.

Luca, Dara Lee, and Michael Luca. 2018. “Survival of the Fittest: The Impact of the Minimum Wage on Firm Exit.” <https://www.hbs.edu/faculty/Pages/item.aspx?num=52552>.

Luca, Michael. 2016. “Reviews, Reputation, and Revenue: The Case of Yelp.com.” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1928601.

Luca, Michael, and Georgios Zervas. 2016. “Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud.” *Management Science* 62(12): 3412–27.

Masterov, Dimitriy V., Uwe F. Mayer, and Steven Tadelis. 2015. “Canary in the e-Commerce Coal Mine: Detecting and Predicting Poor Experiences using Buyer-to-Seller Messages.” In *EC ’15 Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pp. 81–93.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. 2014. “Promotional Reviews: An Empirical Investigation of Online Review Manipulation.” *American Economic Review* 104(8): 2421–55.

Meyer, Robinson. 2014. “Everything We Know About Facebook’s Secret Mood Manipulation Experiment.” *The Atlantic*, June 28.

Miller, Claire Cain. 2011. “U.S. Clears Google Acquisition of Travel Software.” *New York Times*, April 8.

Ostrovsky, Michael, and Michael Schwarz. 2016. “Reserve Prices in Internet Advertising Auctions: A Field Experiment.” <http://web.stanford.edu/~ost/papers/rp.pdf>.

Satariano, Adam, and Jack Nicas. 2018. “E.U. Fines Google \$5.1 Billion in Android Antitrust Case.” *New York Times*, July 18, 2018.

Scott, Mark. 2017. “Google Fined Record \$2.7 Billion in E.U. Antitrust Ruling.” *New York Times*, June 27.

Shapiro Carl, and Hal R. Varian. 1999. “Information Rules: A Strategic Guide to the Network Economy.” *Harvard Business School Press*.

Sun, Monic. 2012. “How Does the Variance of Product Ratings Matter?” *Management Science* 58(4): 696–707.

Varian, Hal R. 2007. “Position Auctions.” *International Journal of Industrial Organization* 25(6): 1163–78.

Verma, Inder. 2014. “Editorial Expression of Concern: Experimental Evidence of Massivescale Emotional Contagion through Social Networks.” *PNAS* 111(29).

Wager, Stefan, and Susan Athey. Forthcoming. “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests.” *Journal of the American Statistical Association* 109(508).

Wu, Lynn, and Erik Brynjolfsson. 2015. “The Future of Prediction: How Google Searches Fore-shadow Housing Prices and Sales.” In *Economic Analysis of the Digital Economy*, edited by Avi Goldfarb, Shane M. Greenstein, and Catherine E. Tucker, 89–118. University of Chicago Press.

Zhu, Feng, Susan Athey and David Lane. 2018. “Apple Pay and Mobile Payments in Australia (A).” *Harvard Business School Case no. 619-010*.

Parag Pathak: Winner of the 2018 Clark Medal

Ariel Pakes and Joel Sobel

Parag Pathak was born in the Finger Lakes region of New York State to parents who had immigrated from Kathmandu, Nepal. Parag’s father obtained a medical degree in Nepal, did an ear-nose-and-throat residency at Georgetown University, and practiced in New York. Parag’s mother went to university in India. She was born into a famous family of priests that dates back to the early nineteenth century. Her grandfather was a noted Sanskrit scholar with the title “Nayaab Bada Guruju,” which translates as “Great Teacher” or “Royal Preceptor.” Parag has two sisters. Rachana is an assistant attorney general in New York. While Parag was still at home, she taught at Elmira College and Parag was often a sounding board for her lectures. Parag credits her with developing his early interest in social science. Sapana is an emergency room physician in Charlotte, North Carolina. Parag’s youth was spent in a scholarly family environment.

Parag received his AB summa cum laude in Applied Mathematics from Harvard in 2002, and a SM in Applied Math the same year. He received his doctorate in economics from Harvard in 2007. Since then, Parag has been employed at MIT where he is presently the Jane Berkowitz Carlton and Dennis William Carlton Professor of Microeconomics. Parag is married to Dr. Rhuma Rjbhandati, and they have two children. The American Economic Association awarded the 2018 John Bates Clark Medal to Parag Pathak for his research on the impacts of educational policies. We will survey some highlights from this work, as listed in Table 1.

■ *Ariel Pakes is the Thomas Professor of Economics, Harvard University, Cambridge, Massachusetts. Joel Sobel is Professor of Economics, University of California—San Diego, La Jolla, California. Their email addresses are apakes@fas.harvard.edu and jsobel@ucsd.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at

<https://doi.org/10.1257/jep.33.1.231>

doi=10.1257/jep.33.1.231



Parag Pathak

An Overview of the Economic Researcher

Parag is part theorist, part empiricist, and fully problem-motivated. His major work is grounded in the institutional details of the educational system. The theory identifies ways to improve that system, while the empirical work determines what aspects of the theory were most important to implement and evaluates the consequences of policy changes. In addition, Parag actively communicates his and his coauthors' findings to the policy community. As a result, those findings have had a major impact on school reforms in many cities both within the United States and abroad. This is a rare combination of qualities for any economist, but especially rare for an economist under the age of 40.

Parag's most influential contributions are about market design and its application to the problem of the allocation of students to schools. That literature dates to a paper by Abdulkadiroğlu and Sönmez (2003) in the *American Economic Review* that motivated the head of the Boston Public School system to contact a group working on mechanism design. New York City, which was using an allocation system that was not functioning well, made a similar request.

Parag and collaborators analyzed the performance of the systems used in Boston and New York and proposed alternatives that were ultimately implemented. Parag's subsequent research on school allocation used theory to deepen our understanding of the implications of both the mechanisms that had not yet been reformed and of the institutional constraints that had been incorporated into some of the reformed allocation mechanisms. These implications were communicated to administrators,

Table 1

Selected Papers by Parag Pathak

-
-
1. “The New York City High School Match” (with Atila Abdulkadiroğlu and Alvin Roth). 2005. *American Economic Review* 95(2): 364–367.
 2. “The Boston Public Schools Match” (with Atila Abdulkadiroğlu, Alvin Roth, and Tayfun Sönmez). 2005. *American Economic Review* 95(2): 368–371.
 3. “Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism” (with Tayfun Sönmez). 2008. *American Economic Review* 98(4): 1636–52.
 4. “Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match” (with Atila Abdulkadiroğlu and Alvin E. Roth). 2009. *American Economic Review* 99(5): 1954–78.
 5. “Incentives and Stability in Large Two-Sided Matching Markets” (with Fuhito Kojima). 2009. *American Economic Review* 99(3): 608–627.
 6. “Inputs and Impacts in Charter Schools: KIPP Lynn” (with Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, and Christopher R. Walters). 2010. *American Economic Review* 100(2): 239–43.
 7. “Accountability and Flexibility in Public Schools: Evidence from Boston’s Charters and Pilots” (with Atila Abdulkadiroğlu, Joshua D. Angrist, Susan M. Dynarski, and Thomas J. Kane). 2011. *Quarterly Journal of Economics* 126(2): 699–748.
 8. “Who Benefits from KIPP?” (with Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, and Christopher R. Walters). 2012. *Journal of Policy Analysis and Management* 31(4): 837–60.
 9. “Explaining Charter School Effectiveness” (with Joshua D. Angrist and Christopher R. Walters). 2013. *American Economic Journal: Applied Economics* 5(4): 1–27.
 10. “School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation” (with Tayfun Sönmez). 2013. *American Economic Review* 103(1): 80–106.
 11. “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools” (with Atila Abdulkadiroğlu and Joshua Angrist). 2014. *Econometrica* 82(1): 137–96.
 12. “The Distributional Consequences of Public School Choice” (with Christopher Avery). 2015. NBER Working Paper 21525.
 13. “The Efficiency of Race-Neutral Alternatives to Race-Based Affirmative Action: Evidence from Chicago’s Exam Schools” (with Glenn Ellison). 2016. NBER Working Paper 22589.
 14. “Charters Without Lotteries: Testing Takeovers in New Orleans and Boston” (with Atila Abdulkadiroğlu, Joshua D. Angrist, and Peter D. Hull). 2016. *American Economic Review* 106(7): 1878–1920.
 15. “The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match” (with Atila Abdulkadiroğlu and Nikhil Agarwal). 2017. *American Economic Review* 107(12): 3635–89.
 16. “Minimizing Justified Envy in School Choice: The Design of New Orleans’ OneApp” (with Atila Abdulkadiroğlu, Yeon-Koo Che, Alvin E. Roth, and Olivier Tercieux). 2017. NBER Working Paper 23265.
 17. “Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism” (with Nikhil Agarwal and Paulo Somaini). 2018. *Econometrica* 86(2): 391–444.
 18. “Reserve Design: Unintended Consequences and the Demise of Boston’s Walk Zones” (with Umur M. Dur, Scott Duke Kominers, and Tayfun Sönmez). 2018. *Journal of Political Economy* 126(6): 2457–79.
-

generating both additional impetus for reforms in the older allocation mechanisms and further refinements in those mechanisms proposed by Pathak and his collaborators. The depth of Parag's knowledge of the institutions in Boston and his ability to communicate ways to improve the assignment mechanism to administrators led the late Mayor Thomas Menino to appoint Pathak as chief technical advisor for Boston's student assignment plan in 2013.

While focusing on school assignment, the knowledge of the institutional structures in the educational system that Parag developed led him and his coauthors to innovative empirical work on the impacts of different types of schools in different neighborhoods. This included empirical results on the impact of charters, pilots, and exam schools.

The data generated by the new allocation mechanisms has enabled Parag and a somewhat different set of coauthors (though not totally different; Parag has collaborated extensively with Atila Abdulkadirođlu on both theoretical and empirical work) to estimate detailed models of school choice that are capable of empirically analyzing a wide assortment of questions. The choice models, in conjunction with different assignment rules, allow for counterfactual analysis that enables a quantitative comparison of different allocation systems. They can also be used to analyze the effect of the assignment mechanisms on students with different characteristics. These papers have proved particularly valuable to administrators, as they clarify just what aspects of the reforms are central to the gains that come from different assignment rules.

The Underlying Issue: Allocation Mechanisms That Do Not Involve Price

Design of a school allocation system is part of a class of problems that arise when objects have to be allocated without the use of prices. In markets where prices are used, allocations are determined by a combination of factors: tastes, the prices of the goods marketed, and endowments. However, there are many circumstances in which society (or a relevant subset like a sports league) does not want to allow endowments to overly impact allocations (the National Basketball Association conducts a draft rather than letting wage offers clear the market for the entering cohort of players). One example is our society's notion that students should have equal opportunity to their preferred schools (at least for the schools in the public school system).

Of course, there is a question of defining what equal opportunity means in this context, but conditional on our definition, we would like to allocate positions in schools in as efficient a way as possible. Families have heterogeneous tastes; they differ in their preferences for school characteristics (importantly in the location of the school, but also in the importance of different educational resources). An allocation mechanism with desirable properties requires a procedure for eliciting those preferences. This is done through the requirement that families submit a preference list for the schools in the system. Consequently it would be useful if the allocation mechanism be "strategy proof"; that is, it should be in the interest of families to report their preferences honestly.

What other properties would we like? The most obvious other desirable property is Pareto efficiency; we should not be able to reallocate and improve one person's welfare without reducing somebody else's welfare. For school choice problems a stability property called "no justifiable envy" is also important. Assume that schools assign a priority to each student. An allocation has justifiable envy if a student would prefer another school to his assignment and a student of lower priority has a place at the preferred school.

The School Matching Problem: Some Background

The goal of the school choice problem is to assign each student to exactly one school. Students submit a ranking of schools to the school administration. Schools have capacities that limit the number of students that they can accept. They also have priorities, which are rankings over students. The priorities reflect institutional restrictions—for example, obligations to serve students who live nearby the school before admitting students who live further away, or to serve students who already have an older sibling in the school. A matching is an assignment of schools to students such that each student is assigned to one school and, for each school, the number of students assigned to the school does not exceed its capacity.

When Pathak began working on this problem, the theoretical literature on matching mostly studied variations on the college assignment problem introduced by Gale and Shapley in a celebrated 1962 paper. Roth and Sotomayor had published a monograph surveying matching theory in 1990. The theory already had significant impact on the design of markets, most prominently in the use of a matching algorithm to assign graduates of medical school to residency programs (Roth and Peranson 1999). The college assignment problem of Gale and Shapley is quite similar to the school choice problem. Students have rankings over colleges, and colleges have rankings over students. Colleges also have capacities that limit the number of students that they can accept. A matching is an assignment of schools to students such that each student is assigned to one college and, for each college, the number of students assigned to the college does not exceed its capacity. Gale and Shapley formulate the problem and then present the "deferred acceptance" algorithm, which is a procedure that generates a matching that has desirable properties. We describe the deferred acceptance algorithm in more detail later.

Abdulkadiroğlu and Sönmez's (2003) paper was the immediate precursor for Pathak's work. Abdulkadiroğlu and Sönmez formulated the school choice problem and compared it to the college assignment problem. They identified the most important formal difference between the problems. In the college assignment problem, each side of the market has preferences defined over the other side (and staying unmatched). Efficient matches have the property that there is no other match that makes all agents and all colleges better off. In school matching, it makes sense to consider the welfare of students, but it is less meaningful to talk about the preferences of schools. A match in the school choice problem is efficient if it is not possible to find another match that makes all students better off.

The literature notes that schools may face constraints about the kind of students they would like. To take account of this, the school matching literature considers allocation mechanisms that are consistent with “priorities”—in particular, the “no justifiable envy” property is conditional on the priorities assigned to students. Priorities may be the same for all schools. For example, if there are district-wide tests, schools may assign higher priority to students with higher test scores. Priorities may be specific to each school. For example, schools may assign higher priority to students who live closer to the school or have an older sibling who attends the school. Priorities may be set by the central administration (as in New York’s system prior to reforms) or by individual schools (as in Boston’s system prior to reforms). Some aspects of priorities may be captured by student preferences, but others will not be. For example, if Student 1 lives close to School 1 and Student 2 lives close to School 2, but Student 1 prefers School 2 to School 1 and Student 2 prefers School 1 to School 2, then the students would be happy with an assignment that places Student 1 in School 2 and Student 2 in School 1. The district, however, might not like the transportation costs associated with this assignment. When students near to the school have higher priority, then the assignment of Student 1 to School 1 and Student 2 to School 2 will not have justifiable envy.

Abdulkadiroğlu and Sönmez (2003) point out that satisfying all three of the desirable properties introduced above—1) strategy-proofness, 2) efficiency, and 3) no justifiable envy—is generally not possible. (This observation is a consequence of a result of Balinski and Sönmez 1999.) Abdulkadiroğlu and Sönmez discuss how the Gale–Shapley deferred acceptance algorithm applies to the school choice problem. The field of market design was emerging from infancy and headed for practical applications.

Parag’s first two papers [1 and 2, as listed in Table 1], published while still a graduate student, appeared in the *Paper and Proceedings* issue of the *AER* in 2005. “The Boston Public School Match” (with Abdulkadiroğlu, Roth, and Sönmez) details the old allocation mechanism in Boston, explains the impact of the absence of incentives for students to report their preferences truthfully in that system, and outlines the way an incentive-compatible system would work in the schooling context. “The New York City High School Match” (with Abdulkadiroğlu and Roth) describes the old New York City high school system and the problems that it created. These articles identified ways in which the existing systems failed to satisfy desirable properties. The authors point to evidence that parents and students were confused about how the mechanisms operated, essentially showing that the mechanism was not strategy-proof and that outcomes were inefficient, leaving students with poorer matches than was possible. These observations explain why New York chose the student-proposing deferred acceptance allocation mechanism that we will describe below.

Subsequently, Parag’s research on school allocation focused on using theory to deepen our understanding of the consequences of different allocation mechanisms given the institutional constraints the schools systems faced. To discuss Parag’s contributions in these matters in more detail, it is useful to describe four mechanisms: serial dictatorship, first preferences first, the student-optimal deferred acceptance mechanism, and the top-trading-cycle mechanism.

In a “serial dictatorship” mechanism, students make choices in sequence. It requires a method, typically based on test scores, which orders students. The first student then picks a favorite school. Subsequent students pick their favorite school subject to availability of seats. Serial Dictatorships are efficient and strategy-proof, but they are guaranteed to lead to stable outcomes only when the priorities assigned to students by every school is equal to the ordering of students. This situation arises in the college matching mechanisms in some countries where students are ordered by outcomes on national entrance exams. Serial dictatorship is inherently unfair because the students who choose early in the sequence have more choices than those who move later. This procedural unfairness influences the final match unless the order students make choices is equal to a common priority of the students.

Prior to Parag’s research, several locations assigned students to schools using some form of a “first preferences first” or the old Boston mechanism. In these mechanisms, students rank schools. Assignments begin by looking at first choices. At each school, students who ranked a particular school first are assigned one-by-one to that school according to the priority assigned by the school until either all the seats at the school are filled or all students who ranked the school first are assigned to the school. If unassigned students remain, the mechanism moves to the next step. The next step looks at the second choice of the students yet to be assigned and assigns these students one-by-one according to priority assigned by the school until no seats remain or all unassigned students who ranked the school are assigned. The k th step of the algorithm looks at the k th choice of all students yet to be assigned and proceeds in an analogous manner. This mechanism is not strategy proof. Students should not rank their first choice first if they perceive that there will not be sufficient capacity for them at that school. The two papers published in the 2005 *AER Papers and Proceedings* discuss this weakness of the first-preference-first assignments and explore some of its ramifications.

In the student-optimal deferred acceptance mechanism, students first apply to their first-choice school. Each school rejects the lowest-ranking students in excess of its capacity, keeping the rest of students, but only temporarily. Students not rejected at this step may be rejected later. In the second round, students rejected in the first round apply to the school next in their ranking. Each school considers these students and students who are temporarily held from the previous step together, and rejects the lowest-ranking students in excess of its capacity, keeping the rest of students temporarily. The k th step of the algorithm looks at the students not yet placed, and proceeds analogously. The deferred acceptance mechanism is strategy proof, and guarantees stability (that is, no justified envy), but not efficiency.

The top-trading-cycle mechanism is a modification of Gale’s top-trading-cycle mechanism introduced in a 1974 paper by Shapley and Scarf. The top-trading-cycle algorithm asks each student to designate their favorite school and each school to designate its highest priority student. Starting with any student, these designations create a chain. The odd elements of the chain are students. The even elements are schools. Following any student is the student’s favorite school; following any school is the school’s highest priority student. Because there are a finite number of students,

any such chain must contain a cycle: a series of designations that begins and ends with the same student. The algorithm assigns students in a cycle to the schools that they prefer most. In the next step of the algorithm, students with assignments are removed; the capacities of schools that received a student are reduced; and once again a cycle is found. This process repeats with the remaining students indicating their favorite school (among those with unused capacity) and schools now pointing to the highest priority unassigned student. The top-trading-cycle mechanism is strategy proof and guarantees efficiency but does not rule out justified envy.

Parag and School Assignment

As noted, prior to Pathak's work, Boston and many other cities used first-preference-first procedures. The fact that these mechanisms were not strategy proof was a primary reason for the reforms. There are a number of arguments that lead to prioritizing strategy-proofness. Students could "game" the Boston mechanism, and, as a result, students who had more information could do better than others; a fact that Parag explores in joint work with Sönmez [3], work we discuss in more detail later. The strategy-proof mechanism eliminates the incentive to game and with it the potential for confusion when submitting preferences. For the same reason, it enables school districts to advise students sincerely that it is in their interest to report preferences honestly.

Also, it is much easier to estimate the distribution of utility functions from submitted preferences that are truthful than from those that are not. It is the distribution of utilities that is required for the analysis of the welfare implications of different assignment mechanisms (or, for that matter, any other rule change). As is illustrated in Parag's empirical work with Abdulkadiroğlu and Agarwal [15], in order to predict the outcome of the Boston mechanism, one needs a model for how students form their rank-order list when preference orderings can be strategic, and that model should take into account the student's perceptions of what other students are likely to do.

However, the argument that the deferred acceptance algorithm is strategy proof depends on the assumption that students can submit preferences of arbitrary length. For practical reasons, school districts impose limits on the number of schools that students may rank (and students may not take the trouble to submit a complete ranking). If students can rank only a small number of schools, even the deferred acceptance algorithm is not strategy proof. Students may wish to exaggerate their preference for a school, fearing that they will not have a high enough priority to be assigned to one of their top choices. It was already known that it is impossible to find a mechanism that is strategy proof, stable, and efficient, but even strategy-proofness alone may also be hard to guarantee in practice. Motivated by these constraints and a need to advise administrators, Parag has shown that deferred acceptance is still less subject to manipulation than other mechanisms.

“School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation” (with Sönmez) [10], proposes a way to rank mechanisms that are not strategy proof. This article connects the theoretical results to events related to school choice mechanisms that occurred in Chicago and England around the time they were writing the paper. The Chicago school district changed their assignment system in 2009, asking 14,000 participants to submit preferences under two different mechanisms. The rationale for the change was a concern that the matches were sensitive to unimportant details (“high-scoring kids were being rejected simply because of the order in which they listed their college prep preferences”). However, the change involved moving from a first-preferences-first mechanism that was vulnerable to manipulation to a deferred acceptance mechanism with finite lists that was also vulnerable to manipulation. Pathak and Sönmez [10] provide a framework that can compare two manipulable mechanisms to identify if one is less manipulable than the other. They show that the old Boston mechanism was the most manipulable mechanism, providing an argument in favor of the change. In England, by a 2007 Act of Parliament, the “first-preferences-first” mechanisms were ruled illegal. Just as in Chicago, new manipulable mechanisms were adopted, but Pathak and Sönmez [10] show again that these mechanisms were less manipulable than their predecessor.

One way to see why the deferred acceptance algorithm is harder to manipulate than other procedures is to note that the others are particularly easy to manipulate. Take the Chicago first-preferences-first mechanism in which students submit lists of finite length greater than one. Assume that there are more students than openings at schools so that at least one student is unassigned. Consider a preference profile in which no student can manipulate the Chicago mechanism. It must be the case that students are assigned to their first choice. To see this, note that otherwise one student must be assigned to a school that is not her first choice. But then some school does not fill all of its openings in the first round of the algorithm. An unassigned student could receive a place at that school by ranking it first. This argument suggests that the only preference profiles that are not subject to manipulation in the Chicago mechanism are quite special. The deferred acceptance algorithm will work well with these profiles too.

The deferred acceptance algorithm with finite lists may not be strategy proof, but it is less subject to manipulation than alternatives. There is another reason why the possibility of manipulation may be limited. Pathak (with Kojima) has written about the performance of matching markets with a large number of participants in the paper “Incentives and Stability in Large Two-Sided Matching Markets” [5]. Real school choice problems have a lot of participants, and there is a general intuition that incentives to manipulate may decrease in large markets. These results are relevant for the school choice literature reviewed above, but are more broadly significant for applications of market design to other situations. Kojima and Pathak’s paper [5] also studies what happens when mechanisms permit schools and students to submit truncated lists of preferences. The paper shows that the fraction of participants with incentives to misrepresent their preferences when others are truthful approaches

zero as the market becomes large. Hence, it provides another reason to be reassured that deferred acceptance works well, even when students submit truncated preference lists. The observation that the Chicago mechanism is subject to manipulation unless all students assigned to schools are assigned to their first choice suggests that large numbers alone will not make other mechanisms less subject to manipulation.

The complexity of certain school choice mechanisms leads to distributional concerns that Parag captures in another paper with Sönmez: “Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism” [3]. The deferred acceptance algorithm identifies a matching that Pareto dominates any equilibrium outcome of the Boston mechanism provided everyone is submitting their true preference ordering. The puzzle is that some parent groups resisted the change from the Boston mechanism to the deferred acceptance mechanism. The paper provides a compelling solution to the puzzle. Pathak and Sönmez look at outcomes of the Boston mechanism when a subset of the population is naïve (and reports preferences honestly) while the rest of the population is strategic. The equilibrium with mixed levels of sophistication may have justifiable envy and, importantly, may lead to an assignment in which the sophisticated agents are better off than they are from the deferred acceptance match. The analysis provides a fairness justification for the deferred acceptance mechanism and explains why changes to school assignment mechanisms—even ones that provide Pareto improvements when all players are sophisticated—need not receive unanimous approval. Hence, the deferred acceptance algorithm “levels the playing field” by eliminating an advantage for strategic sophistication built into the Boston mechanism. As noted above, this was one of the goals of the assignment reforms.

Deferred acceptance algorithms are the most common allocation system recommended by market designers. These mechanisms are attractive because they generate stable outcomes (eliminating justified envy) while maintaining incentives for truthful revelation. They have the theoretical weakness in that they do not provide Pareto-efficient matches. No incentive-compatible mechanism can both eliminate justified envy and guarantee efficiency, but the deferred acceptance mechanism comes close in the sense that it weakly dominates all other incentive-compatible mechanisms that eliminate justified envy. “Minimizing Justified Envy in School Choice: The Design of New Orleans’ OneApp” (with Abdulkadiroğlu, Che, Roth, and Tercieux) [16], establishes a dual result for the top-trading-cycle mechanism. It shows that no incentive-compatible Pareto-efficient mechanism has less justified envy (fewer blocking pairs) than the top-trading-cycle mechanism. Using data from New Orleans (which at the time used a top-trading-cycle mechanism), the paper demonstrates (in a setting not covered by the paper’s theorem) the ability of top-trading-cycle mechanisms to perform better than other procedures that are Pareto-efficient and incentive compatible. The practical message of the paper is a new argument for using top-trading-cycle mechanisms when efficiency is the primary goal.

Pathak has moved from making persuasive arguments that led to changes in allocation systems, to studying details of actual markets to better align theory to practice, to conducting detailed analysis of the performance of school allocation methods.

After schools gained experience with new allocation procedures, there were opportunities to evaluate how the reforms were working. The research brought some new conceptual challenges. Theoretical studies of the college assignment problem recognized that important results required the assumption of strict preferences. For the most part, the literature ignored the problem of a possible tie in the ranking of different preferences. It viewed such ties as unlikely to arise in practice. However, ties are a practical concern in the school assignment problem. For the schools, priorities play the role of preferences. Priorities frequently do not distinguish between pairs of students (two students who both live outside of the school district and have no siblings in the school may have the same priority). One can implement the algorithm by breaking these ties arbitrarily, but tiebreaking rules have consequences. The deferred acceptance algorithm still generates stable and strategy-proof outcomes, but some stability constraints are artificial consequences of tiebreaking rules and may have negative impact on efficiency.

“Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match” (with Abdulkadiroğlu and Roth) [4] is an example of a study that assesses the consequences of reforms in the matching procedures. Because tiebreaking rules have consequences, it is important to study the implication of different kinds of rules. It is useful to distinguish single-tiebreaking (assigning an order to students that breaks ties the same way for all schools) and multiple-tiebreaking (using different tiebreaking rules at different schools). Motivated by simulation results that showed advantages of single tiebreaking rules, the paper demonstrates that although there are outcomes that can be produced using the deferred acceptance algorithm with multiple tiebreaking that cannot be produced using single tiebreaking, these outcomes will not be student-optimal stable matchings. They also show that there is no tiebreaking rule that is strategy proof and dominates deferred acceptance with single tiebreaking. These results acknowledge that ties may cause problems (single tiebreaking is not guaranteed to lead to a student-optimal stable match), but describe a sense in which single tiebreaking provides as good an allocation as alternatives.

Another way in which the practical implementation of matches differs from the theory is the possibility that priorities differ for different subsets of the schools’ seats. In “Reserve Design: Unintended Consequences and the Demise of Boston’s Walk Zones,” Pathak and coauthors Dur, Kominers, and Sönmez [18] identify unusual properties of matching mechanisms when priorities for school seats have a slot-specific nature. For example, in Boston, initially walk-zone priority applied at half of a school’s seats, while it did not at the other half. Students were allowed to apply to both halves, but the order of their application in both had an important effect on the overall assignment. Surprisingly, the fact that the slots were processed sequentially resulted in an assignment nearly identical to that without any walk-zone priority, despite the perception that walk-zone applicants gain an edge. The paper establishes formal results on priorities and precedence, and describes how transparency on these results contributed to the end of Boston’s walk-zone priority.

Parag and Educational Policy

As Parag got to know the administrative structure of different school systems, he became aware of opportunities to use that structure to unravel policy-relevant facts on the impacts of other aspects of educational policy. We focus on his contributions to the understanding of the impacts of charter schools and then note how Parag and his coauthors' research have interfaced with broader aspects of educational policy.

In "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots," Parag and coauthors (Abdulkadiroğlu, Angrist, Dynarski, and Kane) [7] examine the impact of two competing models of school autonomy in Boston on student achievement. Charter schools are treated as their own independent school districts and are not subject to the teachers' union contract. Pilot schools have most of the flexibility of charter schools but continue to be covered by the union contract provisions for the teachers. The authors use the random assignment nature of lotteries for entry into oversubscribed charter and pilot schools in Boston as a plausible identification strategy. They compare test scores of students with similar backgrounds who applied and were not accepted to an oversubscribed school to an accepted student, three years after the lottery decision was made. The results are striking. On one hand, among the students who subscribe to an oversubscribed charter school, winning the lottery is consistently associated with large increases in test scores. On the other hand, among students who subscribe to an oversubscribed pilot school that use lotteries to determine acceptance, winning the lottery is not associated with increased performance.

These results left open two questions. First, what were the characteristics of the oversubscribed charters that led to their effectiveness in serving the population that applied to them? Second, what were the characteristics of the students who applied and benefitted from them? A series of papers pursued these issues. In joint research with Angrist, Dynarski, Kane, and Walters [8], Pathak conducted the first evaluation of a Knowledge is Power Program (KIPP) charter school using assignment lotteries. The KIPP schools are the so-called "No Excuses" schools and feature a long school day and year, selective teacher hiring, strict behavior norms, and encourage a strong student work ethic. They are the largest charter school system in the United States.

The KIPP schools in Lynn, Massachusetts, were initially undersubscribed and then oversubscribed. Using the lottery system in the oversubscribed years to construct a quasi-experimental evaluation, "Who Benefits from KIPP?" [8] and "Inputs and Impacts of Charter Schools: KIPP Lynn" [6] (both with coauthors Angrist, Dynarski, Kane, and Walters) provide evidence that KIPP Lynn generated substantial score gains for lottery winners, with the estimates being remarkably similar to those reported for Boston charters. The gains seemed a bit larger for those who entered with lower achievement levels.

These results added to a growing body of evidence suggesting that urban charter schools have the potential to generate impressive achievement gains, especially for minority students living in high-poverty areas. A puzzling fact is that there

is little evidence of achievement gains at charter schools outside of high-poverty urban areas. In “Explaining Charter School Effectiveness” [9], Pathak and coauthors Angrist and Walters examine a large sample of charter schools throughout Massachusetts using the lottery research design. The paper indicates that the relatively higher effectiveness of urban charter schools might be explained by adherence to the “No Excuses” approach to urban education discussed above.

One problematic feature of the lottery studies is that they rely on select samples of students: specifically, those who apply to a subset of schools and were lottery assigned. Were a more inclusive segment of the population to attend charters, the average effect of charters might be different. In the paper “Charters Without Lotteries: Testing Takeovers in New Orleans and Boston,” Parag and his coauthors Abdulkadiroğlu, Angrist, and Hull [14] look at schools that were taken over by charters, focusing on students who were grandfathered into the charter system. Following Hurricane Katrina, state legislation allowed the Louisiana Department of Education to take control of and delegate the operation of low-performing schools to outside operators. By 2015, the Recovery School District became the first all-charter school district in the United States. Takeovers of underperforming schools have also occurred in Boston and are increasingly being used in other states and countries. Though there is still some selection involved in the takeover experiments (some students may switch out of the district to which the school is assigned) the selection problem in takeover studies is likely to be much less serious than in situations where student apply to charters. Their results from the takeover studies in low-performing urban environments suggest that charters boost achievement by as much or more than the gains estimated from lottery studies in low-performing urban environments.

There are a number of other educational policy areas where Parag and his coauthors have been influential contributors. In “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools” [11], Parag together with Abdulkadiroğlu and Angrist use a regression discontinuity design to examine whether students who scored close to the acceptance line and were accepted to an exam school in New York and Boston did better than those who were close to the acceptance line but rejected. Marginally accepted students show only scattered gains from attending the exam school. This result depends on the characteristics of the marginally accepted applicants, but it does raise two questions. First, do exam schools help, and if so, whom do they help? Second, overall performance at exam schools is much higher than at alternative schools, so the paper also raises questions on the validity of prior findings of peer effects on school achievement. Prior findings were mixed, leaving the possibility that peer effects matter while Parag’s results suggest instead that students perform better because they are more qualified, since the impact of school assignment (and hence of peer achievement) has a small impact on the marginally accepted students.

Parag’s recent work, as yet unpublished, follows up on themes that were related though not central to his prior work but very much in the public policy debate. “The Efficiency of Race-Neutral Alternatives for Race-Based Affirmative Action: Evidence from Chicago’s Exam Schools” (NBER Working Paper 22589) [13] with Glenn Ellison measures the welfare costs of affirmative-action programs. School districts

wish to balance diversity goals with matching high-quality students to high-quality schools. Ellison and Pathak examine admission procedures at elite public schools in Chicago. These schools have shifted from a system that used explicit race-based quotas to one in which schools admit a fraction of their classes on the basis of performance measures only, while allocating the remaining fraction to districts using (not directly race-based) proxies of neighborhood socioeconomic status. The paper makes the straightforward theoretical observation that when racial diversity is valuable, limiting attention to race-neutral schemes is inefficient. It elaborates upon this observation with an analysis of data from the Chicago school district that provides a quantitative measure of the efficiency costs. Diversity goals lead to a reduction in test scores of elite schools, but in the two schools that are the focus of this study, a race-based system would eliminate more than three-quarters of the reduction that seem to be caused by the school district's race-neutral procedure. The paper points out that Chicago's current system also fails to achieve the socioeconomic diversity achieved by a system that takes race into account. Loosely, the efficiency losses arise because a race-neutral system may give priority to low scorers in one district over higher scorers from demographically similar districts.

Chris Avery and Parag present a model in "The Distributional Consequences of Public School Choice" (NBER Working Paper, 21525) [12] to compare school choice to residential-based assignment when housing markets are modeled explicitly. These papers make it clear that Parag has a lot left to contribute to our understanding of the impacts of different education policies, a fact that is likely to maintain interest in Parag's work for some time to come.

Evaluating School Assignment Mechanisms

With more than ten years of experience, it is now possible to evaluate the impact of reforms to school choice mechanisms. In "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match" [15], Parag and coauthors Abdulkadiroğlu and Agarwal return to the original problem of examining school assignment mechanisms, but now with the data and econometric tools that enable them to assess the impact of the reforms. The empirical assessment allows for the impacts of both administrative constraints and possible behavioral differences from what is assumed in the theoretical results. This is a powerful way to assess the impact of the reforms.

This paper and a related paper with Agarwal and Somaini [17] also break new ground methodologically. To evaluate welfare, one must employ a utility function. The fact that applicants provide an ordered ranking of multiple schools enables the authors to use methodology that allows them to estimate an extremely rich set of utility functions quite precisely. Crucial here is the fact that the submitted ranking order across multiple schools generates an ability to let important characteristics, like school and home location, have coefficients that differ across applicants due to unmeasured factors (for example, a parent whose working day is longer than the

school day may submit a ranking with preferred schools all near a relative's home). The second methodologically innovative aspect of these papers is the development of a framework that can evaluate allocations when the submitted preference lists can be strategic—that is, when truth-telling is not a dominant strategy (as in the Boston mechanism). They also show that under certain cognitive assumptions, there is a sense that truth-telling, on the one hand, and the assumption that each agent best responds to the actual distributions of others' play, on the other, bound the results from different allocation mechanisms.

Prior to the reform, New York high school students applied to 5 out of 600 school programs; they could receive multiple offers and be placed on a wait list. The students were allowed to accept only one offer and one wait list, and the process went on for two more rounds. This is labeled the uncoordinated system. After the reform, all schools were integrated into one match, students could rank up to 12 programs, and the student-proposing deferred acceptance algorithm determined the allocation. In both cases, unmatched students were administratively assigned.

The empirical results show a marked improvement in the distribution of welfare as a result of the reform. There were welfare gains for all groups, but they were largest for the disadvantaged. This was mostly because in the new allocation mechanism, fewer students got none of their listed choices. (These students were allocated in the administrative placement round, leading to assignments less desirable than any of the listed choices.) So the main qualitative finding is that the new system matches more than 80 percent of the students in the main round of the algorithm, while prior to reforms this figure was about 50 percent. The students who are assigned administratively do not have a say in their school assignment.

Parag's paper with Abdulkadiroğlu and Agarwal [15] provides comparative results on the following mechanisms: neighborhood assignment, the uncoordinated system, the reformed system, a student-optimal stable match, Pareto-efficient matching, and a utilitarian optimal matching with equal weight to all individuals (this last requires knowledge of the distribution of utility functions, and hence was not administratively feasible). Big increases in welfare were found when going from the neighborhood assignment to the uncoordinated system, and then again from the uncoordinated system to the reformed system that used a deferred acceptance system with an incomplete ranking of schools. However, further increases in going to Pareto efficiency or to a full deferred acceptance algorithm were very small.

These results are extremely useful to administrators as they provide guidance on where to assign priorities when institutional and perhaps cognitive constraints limit their ability to fully implement one of the theoretically preferred mechanisms.

Summary

Parag's work has both improved our understanding of important aspects of the education system and improved the system itself. He combines knowledge of how institutions work with a theorist's ability to formulate models, an applied

economist's ability to develop and use tools for program analysis, and a policy adviser's ability to communicate findings and recommend changes to decisionmakers. He formed working relationships with a wide range of colleagues, each of whom could help him unravel different aspects of the problems that he faced. He created relationships with and acquired detailed knowledge of the workings of many school districts, and the problems they face. This allowed him to do innovative empirical work. From the very start of his career, he has been able to use what he learned on the school districts' problems to motivate new conceptual developments and then to use the theory to solve the motivating problem.

He and his coauthors' work on school admission mechanisms, much of it on the impact of different allocation rules, led to improvements being made to those mechanisms over the last decade and a half. When data became available to empirically analyze the impact of the changes, he developed and applied techniques needed to quantify the impacts. This led to further improvements of the institutions. During the interim, he used the institutional knowledge he had gained from studying admission rules to do innovative studies of other central features of education policy, most notably the impact of charters and other semi-autonomous school systems.

Through all of this Parag has paid great attention to detail, both in the institutions being analyzed and in the analysis per se. This is a major reason that his work has been so influential and is a lesson to economists everywhere. The economics profession has a lot of tools. Parag has mastered many and used them—always with utmost care and precision—to bring fresh insight into the analysis of education policy and ultimately to improve the outcomes of our education system.

■ *We thank Daron Acemoglu, Gordon Hansen, Al Roth, Tayfun Sönmez, and Timothy Taylor for comments.*

References

- Abdulkadiroglu, Atila, and Tayfun Sönmez.** 2003. "School Choice: A Mechanism Design Approach." *American Economic Review* 93(3): 729–47.
- Balinski, Michel, and Tayfun Sönmez.** 1999. "A Tale of Two Mechanisms: Student Placement." *Journal of Economic Theory* 84(1): 73–94.
- Gale, David, and Lloyd S. Shapley.** 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly* 69(1): 9–15.
- Roth, Alvin E., and Elliott Peranson.** 1999. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design." *American Economic Review* 89(4): 748–80.
- Roth, Alvin E., and Marilda A. Sotomayor.** 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monograph Series, Cambridge University Press.
- Shapley, Lloyd, and Herbert Scarf.** 1974. "On Cores and Indivisibility." *Journal of Mathematical Economics* 1(1): 23–37.

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by email at taylort@macalester.edu, or c/o *Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., St. Paul, MN 55105.

Smorgasbord

The committee that awards the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel has published, as is its custom, two essays describing the work of the 2018 winners: William D. Nordhaus and Paul M. Romer. Here's a brief comment on their work from "Popular Science Background: Integrating Nature and Knowledge into Economics." On Nordhaus: "William Nordhaus began his work in the 1970s, after scientists had become increasingly concerned about how the combustion of fossil fuels causes serious global warming, and the detrimental effects of such climate change. Nordhaus took on the daunting task of

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota. He blogs at <http://conversableeconomist.blogspot.com>.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.33.1.247>

doi=10.1257/jep.33.1.247

examining bidirectional feedback loops between human activity and the climate, combining basic theories and empirical results from physics, chemistry, and economics. ... Nordhaus became the first person to design simple, but dynamic and quantitative models of the global economic-climate system, now called *integrated assessment models* (IAMs). His tools allow us to simulate how the economy and climate would co-evolve in the future under alternative assumptions about the workings of nature and the market economy, including relevant policies.” On Romer: “In the early 1980s, when he was a PhD student at the University of Chicago, Paul Romer started developing the theory of *endogenous* growth, where technological advances do not just flow in from external—exogenous—sources, as assumed in earlier economic models. Instead, they are created by purposeful activities in the marketplace. ... He also demonstrated how such endogenous technological change can shape growth, and which policies are necessary for this process to work well. ... Romer believed that a market model for idea creation must allow for the fact that the production of new goods, which are based on ideas, usually has rapidly declining costs: the first blueprint has a large fixed cost, but replication/reproduction has small marginal costs. Such a cost structure requires that firms charge a markup, i.e. setting the price above the marginal cost, so they recoup the initial fixed cost. ... Romer also showed that growth driven by the accumulation of ideas, unlike growth driven by the accumulation of physical capital, does not have to experience decreasing returns. In other words, ideas-driven growth can be sustained over time.” October 8, 2018. The “Popular Science” essay is at <https://www.nobelprize.org/uploads/2018/10/popular-economicsciencesprize2018.pdf>. The more detailed essay, “Scientific Background: Economic growth, technological change, and climate change,” is at <https://www.nobelprize.org/uploads/2018/10/advanced-economicsciencesprize2018.pdf>.

On average, women around the world are paid 15.6 percent per hour less than men according to the *Global Wage Report 2018–2019*, published by the International Labour Organization, which devotes two main chapters to the theme “What lies behind gender pay gaps.” “A gender pay gap measured simply—the so-called ‘raw’ or unadjusted gender pay gap—can arise for a multitude of different reasons, including, among others: differences between female and male educational attainments; lower wages in the sectors and occupations in which women are concentrated; differences between female and male participation rates in part-time and full-time work, which are in turn influenced by women’s role as mothers and their care responsibilities; and discrimination in pay between women and men performing equal work or work of equal value. ... [T]he low labour market participation of women vis-à-vis men is a global phenomenon. Irrespective of income level, in all countries and at any age group, women’s participation rates are always below those of men. ... [F]or most countries, the trend in participation rates for women starts to separate further from that of men at about the age of 25–35 years old, coinciding with the beginning of the period of motherhood. Finally, in only a few of the countries shown here (Armenia, Australia, Mongolia, Philippines, Russian Federation, Ukraine) is there any ‘bounce back’ into the labour market for women. In most other countries, it

seems that motherhood has a long-term effect: once the participation of women declines at around the age of 25–30 years, the proportion of women who stay in (or out) of the labour market across all other age groups thereafter remains constant until approximately retirement age. ... Globally, women are still substantially less likely than men to participate in the labour market. The global gap in labour force participation has been estimated at 27 percentage points, and participation gaps remain particularly wide in the Arab States, northern Africa and southern Asia, in each case exceeding 50 percentage points.” November 26, 2018, https://www.ilo.org/global/publications/books/WCMS_650553/lang-en/index.htm.

Kym Anderson, Giulia Meloni, and Johan Swinnen provide an overview of “Global Alcohol Markets: Evolving Consumption Patterns, Regulations, and Industrial Organizations.” “The global mix of recorded alcohol consumption has changed dramatically over the past half century: Wine’s share of the volume of global alcohol consumption has fallen from 34% to 13% since the early 1960s, while beer’s share has risen from 28% to 36%, and spirits’ share has gone from 38% to 51%. In liters of alcohol per capita, global consumption of wine has halved, while that of beer and spirits has increased by 50%.” “As of 2010–2014, alcohol composed nearly two-thirds of the world’s recorded expenditure on beverages, with the rest being bottled water (8%), carbonated soft drinks (15%), and other soft drinks such as fruit juices (13%).” “In early history, wine and beer consumption was mostly positively perceived from health and food security perspectives. Both wine and beer were safe to drink in moderation because fermentation kills harmful bacteria. ... Beer was also a source of calories. For both reasons, beer was used to pay workers for their labor from Egyptian times to the Middle Ages. Wine too was part of some workers’ remuneration and was included in army rations of some countries right up to World War II.” *Annual Review of Resource Economics*, vol. 10, pp. 105–132 (not freely available online).

Brent R. Moulton discusses “The Measurement of Output, Prices, and Productivity: What’s Changed Since the Boskin Commission?” “I use as my starting point the Final Report of the Advisory Commission to Study the Consumer Price Index, commonly known as the Boskin Commission, which kicked off [in 1996] major efforts to improve core economic statistics ... I conclude that the overall bias of the Consumer Price Index has fallen from about 1.1 percent in 1996 to about 0.85 percent today. Because the CPI and other price indexes are used as deflators in the estimation of productivity, these improvements in the CPI and similar improvements in the Producer Price Index (PPI) have fed directly into reducing bias in the productivity statistics. ... I catalog changes in methodology that have affected real output or prices since January 1997. I also offer three recommendations on ways to renew progress on reducing or eliminating bias in multifactor productivity, GDP growth, and related price indexes. Hutchins Center on Fiscal and Monetary Policy at Brookings, July 2018, at <https://www.brookings.edu/wp-content/uploads/2018/07/Moulton-report-v2.pdf>. For a contemporary discussion of the Boskin Commission, see the six-paper symposium on “Measuring the CPI” in the Winter 1998 issue of this journal.

Collections of Essays

The Becker Friedman Institute at the University of Chicago has published *The Monetary and Fiscal History of Latin America*, with one paper on each of the 11 largest Latin American countries from 1960 to 2016. For example, Diego Restuccia writes on “The Case of Venezuela”: “In the post-war era, Venezuela represents one of the most dramatic growth experiences in the world. Measured as real gross domestic product (GDP) per capita in international dollars, Venezuela attained levels of more than 80% of that of the US by the end of 1960. It has also experienced one of the most dramatic declines, with levels of relative real GDP per capita reaching less than 30% of that of the US nowadays. ... The last period, from 2006 to 2016 deserves special discussion. ... First, there is extreme intervention of the public sector in economic activity through expropriation of private enterprises and government intervention of goods distribution systems. ... Second, this is a period of rising debt, both internal and external, with the internal debt becoming the majority of new debt as external sources of financing have become more limited toward the end of the period. Third, there is a decline in the transparency of debt statistics ... Fourth, there was a partial reform of the Central Bank allowing for the discretionary use of foreign reserves. ... As a consequence of these characteristics, and despite one of the largest oil-price booms in recent history, the government has found it harder to obtain new loans with mounting fiscal deficits, resorting to much more substantial seigniorage. This is a period also in which real GDP per capita and labor productivity are contracting, for example, real GDP per capita ... declined between 2013 to 2016 by 30%.” August 2018, at <https://mafhol.uchicago.edu>.

Jay Shambaugh and Ryan Nunn have edited a six-paper report, *Place-Based Policies for Shared Economic Growth*. As one example, Bradley L. Hardy, Trevon D. Logan, and John Parman discuss “The Historical Role of Race and Policy for Regional Inequality.” “[W]e outline the ways that the spatial distribution of the black population has evolved over time and the ways that spatial distribution has interacted with policy to, at times, reduce and exacerbate levels of inequality. Recognizing the ways that past policies explicitly stymied black economic mobility and how current policies have explicitly or inadvertently done the same provides a basis for understanding how to craft future policies to reduce racial inequalities. Furthermore, recognizing the interconnection of discrimination and the spatial distribution of the black population is important for understanding certain components of regional and spatial inequality. ... Neighborhoods with a significant share of blacks in America’s major cities have lagged white neighborhoods on key socioeconomic indicators since at least the 1970s, including earnings, poverty, educational attainment, and employment. These gaps in neighborhood amenities and neighborhood quality persist into the 2000s.” Hamilton Project at the Brookings Institution, September 2018, <https://www.brookings.edu/multi-chapter-report/place-based-policies-for-shared-economic-growth>.

The *Yale Law Journal* has published a five-paper “Forum: Reflections on the 2017 Tax Act.” Michael J. Graetz contributes the “Foreword—The 2017 Tax Cuts: How Polarized Politics Produced Precarious Policy.” “The Democrats’ complaints about the law’s reduction in the corporate tax rate from 35% to 21% ring hollow. Democrats themselves had long realized that the U.S.’s exceptionally high corporate tax rate in today’s global economy—with highly mobile capital and intellectual property income—invited both U.S. and foreign multinational companies to locate their deductions, especially for interest and royalties, in the United States, and to locate their income in low- or zero-tax countries. This is obviously not a recipe for economic success. Both before and after the legislation, Democrats urged a corporate tax rate of 25% to 28%; meanwhile, Donald Trump asked for a 15% rate. So, even if Democrats had been involved in the legislative process, the 21% rate that we ended up with would be in the realm of a reasonable compromise. ... Congress’s greatest challenge in crafting this tax legislation was figuring out what to do about the international tax rules. ... There were essentially two options: (1) strengthen the source-base taxation of U.S. business activities and allow foreign business earnings of U.S. multinationals to go untaxed; or (2) tax the worldwide business income of U.S. multinationals on a current basis when earned with a credit for all or part of the foreign income taxes imposed on that income. Faced with the choice between these two very different regimes for taxing the foreign income of the U.S. multinationals, Congress chose both. ... No doubt analysts can find provisions to praise and others to lament in this expansive legislation, but we should not overlook its most important shortcoming: its effect on federal deficits and debt.” Vol. 128, October 25, 2018, <https://www.yalelawjournal.org/collection/reflections-2017-tax-act>. These papers complement the two papers by Joel Slemrod and Alan Auerbach in the symposium on “The Tax Cuts and Jobs Act” in the Fall 2018 issue of this journal.

The Review of Keynesian Economics has published a nine-paper symposium on “Milton Friedman’s Presidential Address at 50.” In Robert Solow’s opening essay, “A Theory is a Sometime Thing,” he writes, “My mind kept returning to a famous line of dramatic verse: was this the face that launched a thousand ships? Helen of Troy probably never existed, as Marlowe may not have known. But Milton Friedman’s presidential address did exist, and it launched at least a thousand articles. It may not have burnt the topless towers of Ilium, but it certainly helped lead macroeconomics to its current state of refined irrelevance. The financial crisis and the recession that followed it may have planted some second thoughts, but even that is not certain. A few major failures like those I have registered in this note may not be enough for a considered rejection of Friedman’s doctrine and its various successors. But they are certainly enough to justify intense skepticism, especially among economists, for whom skepticism should be the default mental setting, anyway.” The first two essays, by Solow and by Robert J. Gordon, are freely available online. October 2018, <https://www.elgaronline.com/view/journals/roke/6-4/roke.2018.6.issue-4.xml>. These papers complement the three-paper symposium on “Friedman’s Natural Rate Hypothesis After 50 Years” in the Winter 2018 issue of this journal.

Conversations with Economists

Aaron Steelman interviews Chad Syverson on a variety of issues related to productivity and competition. “[I]f I could invent a machine that made everything we consume now and we didn’t have to work an hour, I would take that. That’s not a bad thing. It does create a distributional issue. ... But inherently, we shouldn’t think of it as a problem.” “An important fact is that the skewness of everything is increasing within industries. Size skewness, or concentration, is going up. Productivity skewness is going up. And earnings skewness is going up. ... Is that technological? Is it policy? Is it a little bit of both? I don’t think we really know the answer. That said, I think it’s less of a mystery now than it was when I started working on this many years ago back in graduate school. ... [T]here’s no doubt productivity is correlated with certain kinds of management practices. ... Is that all of the story? No, I don’t think so. If I had to guess, it’s probably 15 to 25 percent of the story. There’s a lot more going on. I think part of it has to do with firm structure. ... An example I talk about in class a lot is when many mainline carriers in the United States tried to copy Southwest and created little carriers offering low-cost service. For instance, United had Ted and Delta had Song. They failed because they copied a few superficial elements of Southwest’s operations, but there was a lot of underlying stuff that Southwest did differently that they didn’t replicate. I think that presents a more general lesson: You need a lot of pieces working together to get the benefits, and a lot of companies can’t manage to do that. It also typically requires you to continue doing what you have been doing while you are changing your capital and people to do things differently. That’s hard.” *Econ Focus*, Federal Reserve Bank of Richmond, Second Quarter 2018, pp. 22–27, at https://www.richmondfed.org/publications/research/econ_focus/2018/q2/interview.

Robert H. Gertner moderates a discussion on the topic “Should public companies do more than maximize profits? BlackRock cofounder Sue Wagner joins Chicago Booth’s Marianne Bertrand, Robert H. Gertner, and Luigi Zingales to discuss the business of business.” Zingales argues: “If you want to donate a lot of money to your alma mater, you can do it directly through the corporation, or you can distribute the money to shareholders and let the shareholders decide if and how they want to donate it. There is no value destroyed by the donation being made at the shareholder level, and because there is more flexibility in that route—and because I have a different alma mater than many of my fellow shareholders, and we all have different ideas about where our money should go—it is better to push that decision down to the shareholder level rather than doing it at the corporate level. ... However, for most social activities, there are some synergies to decision-making at the corporate level. For example, let’s say I really care about the environment, and I am willing to sacrifice some of my profits to have better management of oil spills. ... It costs much more to manage oil spills at the shareholder level than at the corporate level. So maximizing shareholder value and maximizing shareholder welfare are not the same thing. ... If everyone who cares about the environment doesn’t invest in a particular company, it will be controlled entirely by people who don’t care

about the environment, and they'll run the company in the most environmentally unfriendly way. If you care about the environment, why not create an environmentally friendly index fund that includes oil companies, and then go to shareholder's meetings and vote for board members who care about the environment too?" Marianne Bertrand notes: "[W]e would not want to be in an environment where the CEOs of companies, just because they happen to be the CEOs, are deciding for us as a society, as an electorate, which social objectives we care about and which we don't. We hope that we have a political process in place where the preferences of the electorate about spending on schools or spending on alleviating homelessness would be expressed through the political system, but I think there is a concern that without some guidance as to what social goals companies should be pursuing, especially when those social goals are no longer fully aligned with long-term valuation, we might give corporations too much power." *Chicago Booth Review*, August 28, 2018, <http://review.chicagobooth.edu/finance/2018/article/should-public-companies-do-more-maximize-profits>.

Discussion Starters

Andy Polacek describes "Catastrophe Bonds: A Primer and Retrospective." "In 1992 Hurricane Andrew struck Florida and the Gulf Coast, inflicting \$27 billion in damages, of which \$15.5 billion was covered by insurance. ... It led to the failure of eight insurance companies and pushed others to the brink of insolvency. As a result of the losses suffered during Andrew, insurers reevaluated their risk exposure to coastal areas across the country. Homeowners' insurance prices in coastal communities rose markedly to account for the possibility of significant losses, and many large insurers and reinsurers initially reduced their exposure to catastrophic events in coastal regions. ... [T]he demand for natural-disaster-related insurance by households and businesses meant that new capital had to flow into reinsurance. To increase the available capital, the insurance industry created a new financial instrument called a catastrophe bond. A CAT bond is a security that pays the issuer when a predefined disaster risk is realized, such as a hurricane causing \$500 million in insured losses or an earthquake reaching a magnitude of 7.0 (on the Richter scale). The first CAT bonds were issued in 1997, giving insurers access to broader financial markets and offering institutional investors, such as hedge funds, pension funds, and mutual funds, the opportunity to earn an attractive return on investment uncorrelated with the returns of other financial market instruments in exchange for assuming catastrophe insurance risks." *Chicago Fed Letter*, Federal Reserve Bank of Chicago, 2018, No. 405, <https://www.chicagofed.org/publications/chicago-fed-letter/2018/405>.

Scott Lincicome surveys "The 'Protectionist Moment' That Wasn't: American Views on Trade and Globalization." "In fact, recent public opinion polling uniformly reveals that, first, foreign trade and globalization are generally popular, and in fact more popular today than at any point in recent history; second, a substantial portion

of the American electorate has no strong views on U.S. trade policy or trade agreements; third, and likely due to the previous point, polls on trade fluctuate based on partisanship or the state of the U.S. economy; and, fourth, Americans' views on specific trade policies often shift depending on question wording, especially when the actual costs of protectionism are mentioned. These polling realities puncture the current conventional wisdom on trade and public opinion—in particular, that Americans have turned en masse against trade and globalization ...” *Free Trade Bulletin*, Herbert A. Stiefel Center for Trade Policy Studies at the Cato Institute, November 2, 2018, <https://object.cato.org/sites/cato.org/files/pubs/pdf/ftb-72.pdf>.

Science magazine marked the 50th anniversary of Garrett Hardin's essay on “The Tragedy of the Commons” with a suite of seven short comments in “Tragedy Revisited.” Brad Wible writes in a brief overview: “‘Freedom in a commons brings ruin to all.’ So argued ecologist Garrett Hardin ... Hardin questioned society's ability to manage shared resources and avoid an environmentally and socially calamitous free-for-all. In the 50 years since, the essay has influenced discussions ranging from climate change ... to evolution, from infectious disease to the internet, and has reached far beyond academic literature—but not without criticism. Considerable work, notably by Nobelist Elinor Ostrom, has challenged Hardin, particularly his emphasis on property rights and government regulatory leviathans as solutions. Instead, research has documented contexts, cases, and principles that reflect the ability of groups to collectively govern common resources.” Hardin's original December 13, 1968, essay is available at <http://science.sciencemag.org/content/162/3859/1243>. The seven short comments from December 14, 2018 are at <http://science.sciencemag.org/content/362/6420/1236>.



Advancing Knowledge through Data and Research

HUD User is our nation's premier source of housing research and data. Visit www.huduser.gov to access free research publications and data sets. Browse our eBookstore for our most popular housing research in eBook-friendly formats, and download our free apps to quickly obtain data and information on your mobile device. Subscribe to receive email updates through our eLists and check out our datasets webpage at www.huduser.gov, where you can easily find what you need.



U.S. Department of Housing and Urban Development
Office of Policy Development and Research



ADVANCES IN RESEARCH TRANSPARENCY!

The AEA RCT Registry for randomized controlled trials...

- Dedicated to economics and other social sciences
- Central source for 2,100+ studies from over 120 countries
- Free registration and access to current and completed studies

REGISTER A TRIAL >



Register early to increase the credibility of your results!

For more information, go to
<https://www.socialscienceregistry.org/>

NEW!

ECON spark

AEA Discussion Forum



Join the conversation!

Help the AEA build a forum that enriches our community with meaningful, thought-provoking and impactful discussion.



www.aeaweb.org/econspark

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary access to *JEP* articles, go to the AEA website: <http://www.aeaweb.org>. Annual dues for regular membership are \$22.00, \$32.00, or \$42.00, depending on income; for an additional fee, you can receive this journal, or any of the Association's journals, in print. Change of address notice must be received at least six weeks prior to the publication month.

Copyright © 2019 by the American Economic Association. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation, including the name of the author. Copyrights for components of this work owned by others than AEA must be honored. Abstracting with credit is permitted. The author has the right to republish, post on servers, redistribute to lists, and use any component of this work in other works. For others to do so requires prior specific permission and/or a fee. Permissions may be requested from the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203; e-mail: aeainfo@vanderbilt.edu.

Founded in 1885

EXECUTIVE COMMITTEE

Elected Officers and Members

President

BEN S. BERNANKE, The Brookings Institution

President-elect

JANET L. YELLEN, The Brookings Institution

Vice Presidents

LAWRENCE F. KATZ, Harvard University

DAVID H. ROMER, University of California, Berkeley

Members

NICHOLAS BLOOM, Stanford University

ERICA FIELD, Duke University

ADRIANA LLERAS-MUNEY, University of California, Los Angeles

BETSEY STEVENSON, University of Michigan

MARTHA BAILEY, University of Michigan

SUSANTO BASU, Boston College

Ex Officio Members

ALVIN E. ROTH, Stanford University

OLIVIER BLANCHARD, Peterson Institute for International Economics

Appointed Members

Editor, The American Economic Review

ESTHER DUFLO, Massachusetts Institute of Technology

Editor, The American Economic Review: Insights

AMY FINKELSTEIN, Massachusetts Institute of Technology

Editor, The Journal of Economic Literature

STEVEN N. DURLAUF, University of Chicago

Editor, The Journal of Economic Perspectives

ENRICO MORETTI, University of California at Berkeley

Editor, American Economic Journal: Applied Economics

ALEXANDRE MAS, Princeton University

Editor, American Economic Journal: Economic Policy

MATTHEW D. SHAPIRO, University of Michigan

Editor, American Economic Journal: Macroeconomics

SIMON GILCHRIST, New York University

Editor, American Economic Journal: Microeconomics

JOHANNES HÖRNER, Yale University

Secretary-Treasurer

PETER L. ROUSSEAU, Vanderbilt University

OTHER OFFICERS

Editor, Resources for Economists

WILLIAM GOFFE, Pennsylvania State University

Director of AEA Publication Services

JANE EMILY VOROS, Pittsburgh

Managing Director of EconLit Product Design and Content

STEVEN L. HUSTED, University of Pittsburgh

Counsel

TERRY CALVANI, Freshfields Bruckhaus Deringer LLP
Washington, DC

ADMINISTRATORS

Director of Finance and Administration

BARBARA H. FISER

Convention Manager

GWYN LOFTIS



The Journal of
Economic Perspectives

Winter 2019, Volume 33, Number 1

Symposia

Women in Economics

Shelly Lundberg and Jenna Stearns, “Women in Economics: Stalled Progress”

Leah Boustan and Andrew Langan, “Variation in Women’s Success across
PhD Programs in Economics”

Kasey Buckles, “Fixing the Leaky Pipeline: Strategies for Making Economics Work
for Women at Every Stage”

Financial Stability Regulation

Daniel K. Tarullo, “Financial Regulation: Still Unsettled a Decade After the Crisis”

Darrell Duffie, “Prone to Fail: The Pre-Crisis Financial System”

David Aikman, Jonathan Bridges, Anil Kashyap, and Caspar Siegert, “Would
Macroprudential Regulation Have Prevented the Last Crisis?”

Public Provision of Economic Data

Ellen Hughes-Cromwick and Julia Coronado, “The Value of US Government Data to
US Business Decisions”

Hugh Rockoff, “On the Controversies Behind the Origins of the Federal
Economic Statistics”

Ron S. Jarmin, “Evolving Measurement for an Evolving Economy:
Thoughts on 21st Century US Economic Statistics”

Articles

Spencer Banzhaf, Lala Ma, and Christopher Timmins, “Environmental Justice:
The Economics of Race, Place, and Pollution”

Susan Athey and Michael Luca, “Economists (and Economics) in Tech Companies”

Ariel Pakes and Joel Sobel, “Parag Pathak: Winner of the 2018 Clark Medal”

Recommendations for Further Reading

