The Journal of Economic Perspectives

A journal of the
American Economic Association

Fall 2026

The Journal of
Economic Perspectives

A journal of the American Economic Association

Editor
Heidi Williams, Dartmouth College

Coeditors
Jeffrey Kling, Washington, DC, USA
Jonathan Parker, Massachusetts Institute of Technology

Associate Editors
Panle Jia Barwick, University of Wisconsin–Madison
Marianne Bertrand, University of Chicago
Karen Clay, Carnegie Mellon University
Jeffrey Clemens, University of California, San Diego
Michael Clemens, George Mason University
Steven J. Davis, Stanford University
Benjamin Golub, Northwestern University
Andrew Greenland, North Carolina State University
Louis Kaplow, Harvard University
Imran Rasul, University College London
Jeffrey Wooldridge, Michigan State University

Data Editor
Lars Vilhuber

Managing Editor
Timothy Taylor

Assistant Managing Editor
Emma Bender

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

The Journal of Economic Perspectives gratefully acknowledges the support of Macalester College.
Registered in the US Patent and Trademark Office (®).
Copyright © 2026 by the American Economic Association; All Rights Reserved.
Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.
Printed at Sheridan Press, Hanover, Pennsylvania, 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), Fall 2026, Vol. 40, No. 4. The JEP is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. 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 USA.

The Journal of
Economic Perspectives

Contents

Volume 40 • Number 4 • Fall 2026

Symposia

Economics of Higher Education
Zachary Bleemer and Jesse Rothstein, "The Meritocratic Consensus and Stratification in Higher Education" 3
E. Jason Baron, Shawn Kantor, and Alexander Whalley, "Can Government Support for Innovation Create the Next Silicon Valley?" 31
William G. Howell, "Trust Violations in Research Universities" 53

Private Debt
Victoria Ivashina, "The Role of Private Debt in the Financial Ecosystem" 73
Gregor Matvos, Tomasz Piskorski, and Amit Seru, "Private Credit, Balance Sheets, and Financial Stability" 95

Articles
Susan Athey, Neil Gandal, and Tyler Moore, "Systemic Cyber Risk: Linking Measurement, Market Incentives, and Policy" 115
Isaiah Andrews, Sarah Moon, and Jesse M. Shapiro, "In Praise of the Bayesian Bootstrap" 137
Dean Karlan, "The Why and How of Foreign Aid" 159
David J. Deming, Raffaella Sadun, and Christopher Stanton, "The Role of Organization in Firm-Sponsored Training" 185

Feature
Timothy Taylor, "Recommendations for Further Reading" 207

Statement of Purpose

The *Journal of Economic Perspectives* aims to bridge the gap between the general interest business and financial press and standard academic journals of economics. 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

Susan Athey, Stanford University
Cecilia Conrad, Lever for Change
Wenxin Du, Harvard University
Ted Gayer, Niskanen Center
Gary Hoover, Tulane University
Glenn Hubbard, Columbia University
Jeffrey Kling, Congressional Budget Office
Dylan Matthews, Vox.com
Catherine Rampell, Washington Post
Otis Reid, Open Philanthropy
Andrés Rodríguez-Clare, University of California, Berkeley
Ted Rosenbaum, Federal Trade Commission
Sarah West, Macalester College

Journal of Economic Perspectives—Volume 40, Number 4—Fall 2026—Pages 3–30

The Meritocratic Consensus and Stratification in Higher Education

Zachary Bleemer and Jesse Rothstein

The modern American higher education ecosystem came into being in the period following World War II. Increased federal scientific research funding made some institutions much more research intensive, with dramatic increases in revenues and scientific activity. Simultaneously, the college-going rate rose substantially, and states expanded and created a large number of teaching-focused public colleges and universities. An implicit question arose: Which students would attend which institutions?

Most states arrived quickly at the same answer: Students with greater pre-college academic achievement (using traditional metrics like test scores and grade point averages) would be grouped together at better-funded and higher-quality colleges and universities. Students with weaker pre-college achievement would also be grouped, but at lower-funded, teaching-focused institutions. We characterize this idea, taken for granted as the organizing principle of US higher education, as the "meritocratic consensus." In this essay, we review the meritocratic consensus and consider potential justifications for it.

First, we demonstrate that higher education is highly stratified. There are substantial disparities across institutions in expenditures per student, and high-testing students—who tend to be from families with high socioeconomic status—are concentrated at the higher-spending institutions.

■ *Zachary Bleemer is Assistant Professor of Economics, Princeton University, Princeton, New Jersey. Jesse Rothstein is Professor of Public Policy and Economics and David Gardner Chair of Higher Education, University of California at Berkeley, Berkeley, California. The authors are listed in alphabetical order.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251487.

Second, we provide a short history of how this has evolved over the last 125 years. As late as 1960, most US higher education institutions remained effectively open access and variation in per-student expenditures was relatively low. Since 1960, stratification has steadily intensified. In 1969, the top quarter of institutions by selectivity (enrolling the highest-testing students) had per-student instructional expenditures about double those at the institutions in the bottom quarter of test scores. By 1993, this had risen to about 2.5 times and, by 2019, to 3 times.

Third, we discuss some exceptions to the "meritocratic" aspect of the consensus. While high-resource institutions conduct selective admissions that select heavily on academic achievement, they also consider nonacademic criteria like athletic talent, potential donations, and various dimensions of student diversity. These objectives are often described as dimensions of "merit," but they tend to cut against the admission of the most *academically* accomplished students. We argue that relatively few students are shifted toward higher-expenditure institutions by policies like athletic, donor, and legacy preferences. While race-based affirmative action was once a larger factor, it never accounted for more than a few percent of university admissions and was banned outright by the 2023 US Supreme Court decision in *Students for Fair Admissions* (*SFFA*) *v. Harvard* (600 US 181 [2023]).

With this background in place, we discuss three potential economic justifications for the meritocratic consensus. There are clear reasons why one might want to group students together by prior achievement, so we focus on the strong correlation between student achievement and institutional resources. (As we make clear, this is not simply a market outcome, and it appears in the large public sector as well as the smaller private sector.) We consider several reasons that it could be socially desirable to spend more money on the instruction of high-achieving students than on that of students with worse high school achievement. One is that pre-college academic preparation and college resources may be q-complements in educational production—that is, higher-achieving students may benefit relatively more from smaller classes and access to cutting-edge research. A second is that there may be large social value in providing the highest-quality training to students who are most likely to become scientists, successful entrepreneurs, creative geniuses, and others who can expand the boundaries of human knowledge and creativity. Third, meritocratic selection may create useful incentives for high school students to invest in skill development. While all three models are plausible, we find that there is little or no empirical evidence in support of any of them, and some evidence pointing against them.

In the conclusion, we discuss recent developments and the future of the meritocratic consensus. After several decades in which the association between selectivity and per-student university expenditures steadily strengthened, stratification across universities has leveled out since 2010. During this same period, however, there has been a rise in stratification *inside* colleges and universities, particularly at public institutions. These institutions have increasingly created "honors colleges" and other programs designed to separate and provide additional resources to their highest-achieving students. In addition, more than 40 percent of lucrative majors at

Table 1
Characteristics of Four-Year Colleges by Tier

Tier	Share of enr. (percent)	Admit rate (percent)	Avg. SAT	Mean parental inc. rank	Expenditures per student ($)			Students per faculty	Grad. rate (percent)
					Inst.	Res.	Admin.		
Ivy League	0.9	7.4	1502	80.5	68,800	38,300	81,700	10.0	96.1
Other top-50 private	3.3	15.2	1466	80.0	61,200	32,300	41,800	10.4	92.0
State flagship	15.2	64.9	1275	71.7	22,000	15,000	14,800	15.9	78.7
Other public R1	15.2	63.9	1232	68.4	19,800	11,700	14,400	19.5	72.5
Less-sel. private	27.4	67.1	1168	68.5	16,400	1,400	19,800	22.9	64.6
Less-sel. public	36.9	73.1	1105	62.3	12,000	1,700	9,700	23.6	54.6
For profit	1.2	70.6	–	44.3	7,800	0	15,900	141.5	40.3

Source: IPEDS and Chetty et al. (2020).
Note: See the text for college tier definitions, which partition four-year institutions. For each tier, the table reports the share of total full-time, first-year enrollment in four-year institutions and several enrollment-weighted means: the undergraduate admissions rate; SAT score (except for for-profits, which generally do not report average scores); parental income rank; per-student instructional, research, and other administrative expenditures; student-to-faculty ratio; and six-year graduation rate. See Supplemental Appendix A for variable definitions and details on data construction. Six-year graduation rate is measured for the 2017 first-year cohort, parental income rank is measured for the 1989–1991 birth cohort by students' modal enrollment institution, and all other statistics are measured in 2019.

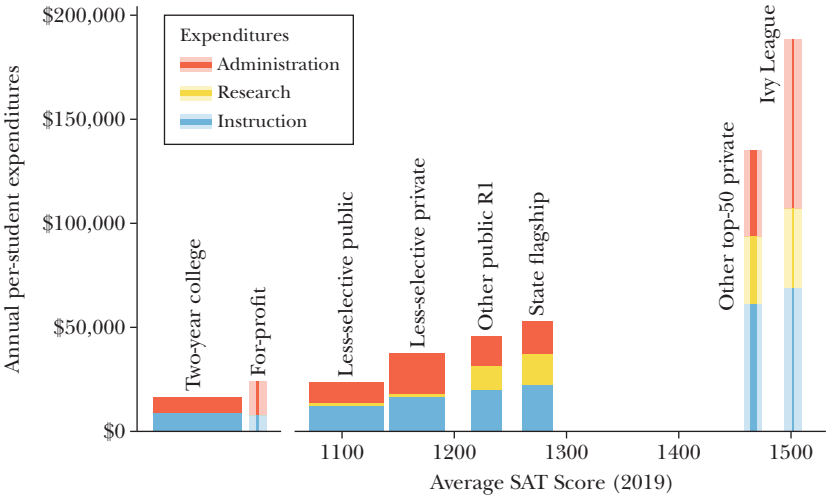
public research universities—like engineering and business—have adopted supplemental achievement-based admissions policies over the past 50 years (Bleemer and Mehta 2024). Because these majors tend to enroll the highest-achieving students from the most advantaged backgrounds, increasing stratification between college majors parallels the across-institution stratification that we document.

The Heterogeneity of Higher Education

About three-fifths of students who graduate high school enroll directly in college the following fall. While not all of those students ultimately graduate, most get some kind of credential: More than half of Americans aged 25–29 have at least a two-year associate degree, while 40 percent have a four-year bachelor's degree or more (de Brey, Snyder, Zhang, and Dillow 2021).

Students receive these credentials from institutions that differ widely. Table 1 shows summary statistics for seven tiers of four-year colleges and universities: the eight Ivy League institutions, the other 42 top-50 private institutions (with a twenty-fifth percentile SAT score in 2019 of 1340 or higher), the 50 state flagship public universities, the 56 nonflagship "R1" public universities classified as having very high research activity (Carnegie Foundation for the Advancement of Teaching 2000), other four-year private and public universities, and for-profit institutions. The top-50 private universities with the highest average SAT scores together enroll only about 4 percent of four-year college students. In contrast, public R1 universities, including "flagships" like the Pennsylvania State University,

Figure 1
Universities' 2019 Expenditures by Tier



Source: IPEDS.
Note: This figure shows average 2019 expenditures per full-time equivalent (FTE) student (counting each part-time student as one-third of an FTE) by tier of institution and type of expenditure. See Supplemental Appendix A for variable definitions. The width of each solid bar is proportional to full-time, first-year enrollment in the tier; semitransparent bars are widened relative to that for visibility. Average SAT scores are not available for two-year and for-profit colleges, where many students do not take admissions tests; these tiers are shown here separately for comparison. All dollar amounts are adjusted using the Consumer Price Index to 2025.

the University of Wisconsin, and the University of Virginia, as well as others like Utah State University and the University of Pittsburgh, enroll about 30 percent of students. About 64 percent of students attend less-selective public or private institutions. The for-profit share of all first-time, first-year college enrollment (including two-year schools not shown in Table 1) was 2.6 percent in 2019, though it was much higher—nearly 10 percent—in the late 2000s.

These tiers of higher education institutions also differ in the educational experiences they offer. The more selective institutions have much higher expenditures, fewer students per faculty member, and higher graduation rates.

Figure 1 shows average instructional, research, and administrative expenditures per pupil at institutions in each tier. These estimates follow university reporting guidelines and are, admittedly, imperfect. For example, faculty salaries are counted as part of "instruction," but of course, some of that amount is payment for doing research.[1] Similarly, libraries and student services may contribute to instruction but are included

[1] Research time may contribute to instruction if it is a compensating amenity for teachers or if it is complementary with teaching.

here as administrative expenditures. However, such issues will not overshadow the enormous differences shown here: The most selective schools spend many times as much as less-selective and two-year colleges, both overall and in dedicated instructional expenditures.

This is the "meritocratic consensus": Selective schools have much higher spending and revenues per student. In 2019, universities' per-student instructional expenditures were 29 percent higher per 100 average SAT points of their enrolled students. The pattern is similar among public and private universities; at public universities, for example, per-student instructional expenditures were 24 percent higher for each 100 average SAT points. Across states, the gradient of per-student instructional expenditures with respect to SAT scores ranges from 10 percent in Illinois and Mississippi to 50 percent in West Virginia and Washington.[2]

Government educational appropriations—including both direct appropriations, which are largely for research purposes, and financial aid programs, like Pell and state grant programs—are roughly evenly distributed across more- and less-selective institutions. The larger expenditures at selective institutions are financed by substantially higher net tuition funding (gross tuition payments minus gross financial aid), private funding (including alumni donations, affiliated groups like booster clubs, and endowment returns), and government research contracts (see Supplemental Appendix Figure A-1).

The institutions that enroll higher-testing students also tend to enroll students from higher socioeconomic backgrounds, as shown in Table 1, which shows average parental income ranks for students in each tier. Figure 2 provides more detail, dividing students into five quintiles by family income and separating the top 1 percent from the rest of the top quintile. About two-thirds of students at Ivy League and other top-50 private universities—and more than half of students at state flagships—come from families in the top income quintile. The income distributions of students at less-selective public and two-year colleges are closer to those of the country as a whole. More than half of students who do not go to college come from families in the bottom 40 percent of the income distribution.[3]

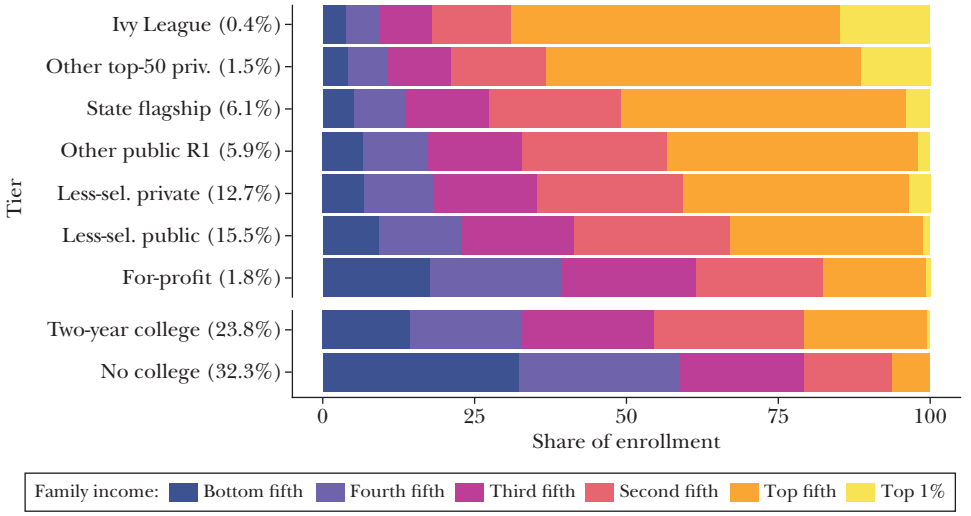
Many researchers have investigated the relationship between university selectivity and value added (Dale and Krueger 2002; Hoekstra 2009; Cohodes and Goodman 2014; Bleemer 2022). Results have been mixed, without clear evidence of a strong relationship. A smaller literature examines the relationship between university expenditures and value added to students. This is clearer, showing a consistent positive association (Chetty, Deming, and Friedman 2026; Bleemer 2021; Mountjoy and Hickman 2021).[4] Deming and Walters (2017) provide suggestive evidence that increasing university expenditures improves students' educational outcomes,

[2] See Supplemental Appendix Figures A-3, A-4, and A-5.

[3] This figure and much of our discussion below exclude nonresident international students at US universities, who accounted for about 3 percent of first-time, full-time undergraduates in 2019.

[4] Of particular note is figure 5 in Mountjoy and Hickman (2021), which reproduces the Dale and Krueger (2002) finding of no relationship between value added and selectivity but finds a positive relationship with institutional resources. Several studies show that decreased public university revenues lead to worse

Figure 2
College Enrollment by Tier and Family Income



Source: IPEDS and Chetty et al. (2020).
Note: This figure shows the parental income distribution of students enrolled at colleges in each of the tiers used in Table 1. Tiers are augmented by additional categories representing two-year colleges (including public, private, and for-profit institutions) and students who do not go to college after high school (including those who do not graduate high school). Tier labels include (in parentheses) the ratio of total US-resident first-time, first-year enrollment in 2000 to the number of high school graduates in 2000 (from de Bray, Snyder, Zhang, and Dillow 2021); the "no college" share is the residual. Parental income is measured among the 1980–1982 birth cohorts, assigning children to their modal institution and parents to quintiles (splitting out the top 1 percent from the top quintile). See Supplemental Appendix A for variable definitions.

suggesting that the correlation is causal and paralleling evidence that increases in K-12 education spending raise student test scores and improve longer-run outcomes (Jackson and Mackevicius 2024; Lafortune, Rothstein, and Schanzenbach 2018; Rothstein and Schanzenbach 2022). If this result is correct, the patterns seen here suggest a pattern of self-reinforcing economic inequality across generations: High-income families provide greater support for their children to achieve greater traditional pre-college academic accomplishment ("merit"), leading those children to enroll at the highest-resource institutions, and thereby conveying an important relative advantage in lifetime income (Rothstein, forthcoming).

quality and lower graduation rates (Kane, Orszag, and Apostolov 2005; Bound and Turner 2007; Bound, Lovenheim, and Turner 2010).

The Rise of the Meritocratic Consensus

Where did the meritocratic consensus come from? This section traces its slow evolution in American higher education over the past 125 years.[5]

The Early Landscape: 1900–1945

At the beginning of the twentieth century, between 5 and 10 percent of young people went to college (de Brey, Snyder, Zhang, and Dillow 2021). Three-quarters of college students were at private institutions, with 10 percent attending the colleges that would soon become known as the "Ivy League" (Goldin and Katz 1999; Bleemer and Quincy 2025). College admission—even at the Ivy League—was generally available to any student who had completed the requisite number of subject-specific high school units. The influential 1892 Committee of Ten—which included Harvard University president Charles Eliot and US Commissioner of Education William Harris among others—recommended that "any successful graduate of a good secondary school should be free to present himself at the gates of the college or scientific school of his choice" (Committee of Ten 1892). Public universities in many states directly accredited their states' high school curricula as a way of enforcing minimum pre-college academic preparation.

Some universities additionally "administer[ed] admission examinations to weed out the clearly incompetent," but this was used only to set a floor, not to conduct competitive admissions. At Harvard, for example, "the size of the Freshman Class was determined … by the number of qualified applicants" (Bender 1952). Neither high school grades nor standardized tests were regularly considered in public or private university admissions (Hinckley 1941), and public universities were, in the 1907 words of University of Illinois President Andrew Draper, oriented toward "put[ting] away all exclusiveness and dedicat[ing] themselves to universal public service" (Veysey 1965). This openness was also manifested in prices: Average net tuition and fees at public and private universities in 1923 were $1,390 and $3,150, respectively, in 2025 dollars (Bleemer and Quincy 2025).

Universities maintained academic standards by evaluating classroom performance rather than standardized test scores; many low-performing students were dismissed. This often resulted in graduation rates below 50 percent at even the most prestigious institutions like Harvard University and Brown University (Thelin 2010).

There was already substantial between-university variation in annual per-student instructional expenditures in the early twentieth century, but far less than in subsequent years. Table 2 shows how cross-university variation in per-student institutional expenditures has risen over the past 100 years.[6] Panel A shows average

[5] We generally rely on enrolled students' average SAT scores (or ACT scores converted to the SAT scale) as our measure of institutional selectivity due to limited availability of high school GPA data. Supplemental Appendix Figure A-7 shows that the SAT rankings of institutions have been largely stable since at least the mid-1980s, and we sometimes use the post-1980 SAT as a proxy for hard-to-measure earlier ordinal selectivity.

[6] We use a smaller number of categories than in earlier tabulations to permit comparability over time.

Table 2
Distribution of Instructional Expenditures per Student over Time

	1929	1969	1984	1996	2008	2019
Panel A. Average instructional expenditures per student (in 2025 dollars)						
All four-year institutions	4,100	8,500	10,700	13,400	14,700	16,800
All public four-year institutions	4,400	8,100	10,800	12,300	14,100	15,800
Ivy League	8,300	23,100	29,500	44,200	77,200	75,000
State flagship	5,000	10,800	12,900	15,800	20,200	21,900
Public nonflagship	3,800	7,700	10,300	11,500	12,600	14,200
Ivy versus flagship gap (%)	68	115	128	181	281	242
Flagship versus nonflagship gap (%)	21	35	35	43	54	50
Panel B. Coefficient of variation of per-student instructional expenditures (%)						
Public institutions	38	30	49	55	50	49
All four-year institutions	53	42	50	59	60	60

Source: Biennial Survey of Education, 1928–1930 (Cooper 1931), HEGIS, and IPEDS.
Note: Means and standard deviations are weighted by enrollment. All series are winsorized at 5 percent to limit the contribution of any single university. Enrollment is measured as total enrollment in 1929 and full-time equivalent enrollment in other years (counting part-time enrollees as one-third). The flagship/nonflagship gap is measured by year-specific regressions of log expenditures on a flagship indicator, including state fixed effects to isolate within-state comparisons. See Supplemental Appendix A for variable definitions.

instructional expenditures per student (converted to 2025 dollars), while panel B shows the coefficient of variation (that is, the standard deviation divided by the mean). At the 111 public institutions that awarded bachelor's degrees in 1929, average per-student expenditures were about $4,400 and the standard deviation of per-student instructional expenditures was about 38 percent of that, or $1,700 (in 2025 dollars). Flagship public universities spent about 21 percent more per student than nonflagships in the same state, while the soon-to-be-named "Ivy League" spent 68 percent more than the flagship publics.[7]

Growing Enrollment and Selectivity: 1945–1990

The share of young men who earned college degrees approximately doubled (from 5 to 10 percent) between 1905 and 1940; it then doubled again in the next decade (Goldin, Katz, and Kuziemko 2006; Stanley 2003). This shock in student demand forced supply-constrained universities to implement novel rationing strategies, including new admissions reviews that often relied on standardized test scores (Karabel 2005; Lemann 1999). States also opened or expanded many public colleges—or new branches of existing colleges—to relieve the pressure on the flagship institutions.

Rutgers, The State University of New Jersey, can serve as a paradigmatic example. Rutgers was an early adopter of standardized testing. In 1939, any student who

[7] Supplemental Appendix Table A-1 shows that gaps in *total* per-student revenues were similar but slightly smaller in magnitude.

satisfactorily completed the curriculum of any high school in New Jersey (or many schools in neighboring states) *or* achieved a minimum standardized test score was admitted to the Rutgers New Brunswick campus (Meder 1939). In fall 1946, Rutgers implemented a new admissions policy following "unprecedented demand for admission": "other things being equal, preference is given to men having the best high school and aptitude test records" (Clothier 1946). Qualified but lower-performing applicants were offered admission to evening courses at the university, or eventually to the new Rutgers campuses at Newark and Camden (created in 1946 and 1950, respectively).

While these new "meritocratic" admissions policies were often intended as temporary responses to a short-run demand shock, the continued growth in college demand motivated their permanent codification. One popular blueprint was the 1960 California Master Plan for Higher Education (henceforth, the Master Plan). The Master Plan laid out a three-tiered higher education system—the research universities of the University of California; the teaching-oriented, bachelor's-only California State University campuses; and two-year community colleges—and suggested using applicants' pre-college academic achievement to track students between the segments. The admissions criteria were valued for their predictive "validity" in identifying students who were likely to successfully complete their studies (Spindt et al. 1961).[8] Each tier massively expanded midcentury; the University of California campuses at Davis, Irvine, Riverside, San Diego, San Francisco, Santa Barbara, and Santa Cruz were all established between 1954 and 1965, in some cases by converting existing institutions, and eight new California State University campuses were opened. The authors of the Master Plan nevertheless recognized that rising demand among qualified students for certain institutions—especially the University of California, Berkeley, and the University of California, Los Angeles—would soon outstrip capacity, and suggested two remedies: (1) campus-level selective admissions policies designed such that only "the best students should be granted their first choice," which were implemented for the first time in California public universities in 1974 (Karabel 1989); and (2) diverting demand to other campuses, via "increased prestige of the junior colleges" and "more personal instruction, a richer student life, and superior housing and parking facilities" at the less-oversubscribed University of California and California State University campuses (Coons et al. 1960). The latter has proven difficult to achieve.

Fewer than 5 percent of higher education institutions relied upon tests for admission, but this rose steadily over the subsequent 50 years. Private institutions led the way, in part because their greater geographic market integration made standardized measures of pre-college academic achievement more valuable to evaluate students from distant and unfamiliar high schools (in this journal, Hoxby 2009). By 1962, over 50 percent of private universities, but only 28 percent of public

[8] Three studies on the predictive validity of College Board standardized tests were published in 1960. By 1970, over 30 scholarly publications per year were evaluating the SAT's "predictive validity" (as indexed by Google Scholar); by 2010, this number had risen above 150 (for example, Westrick et al. 2019).

institutions, required tests in admissions. By the early 1980s, almost all private colleges and universities required test scores, and public institutions had caught up by the end of that decade (Supplemental Appendix Figure A-6).[9]

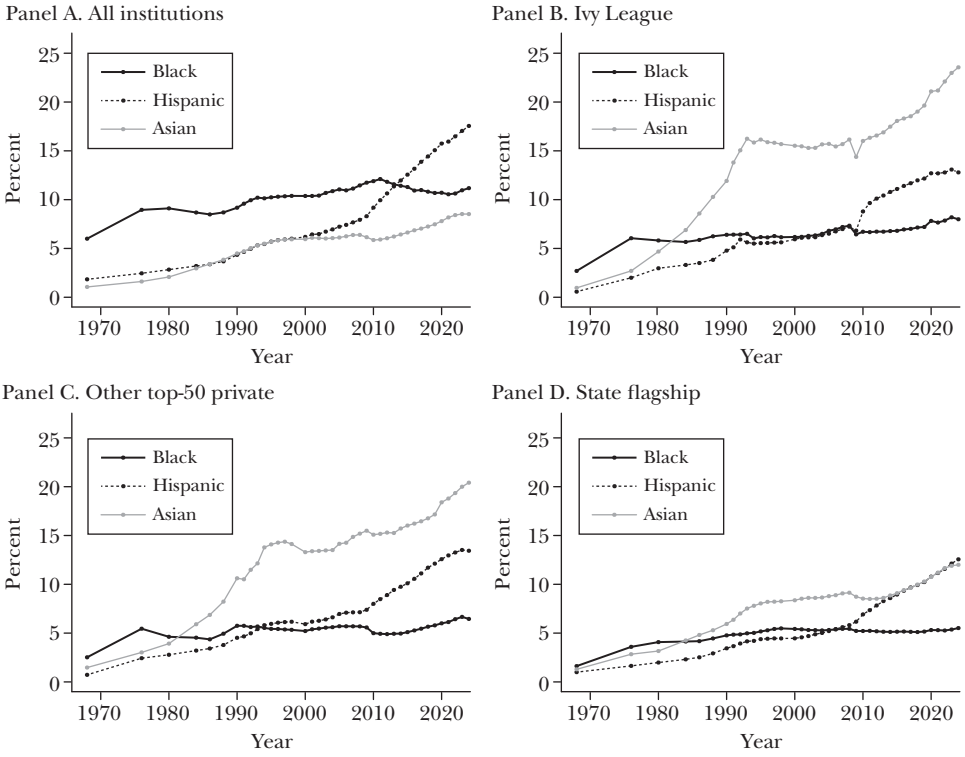
The use of standardized tests in admissions, coupled with strong student preferences that are correlated with university resources, results in strong assortative matching between high-resourced universities and high-scoring students. Students' preferences could arise either because students have direct preferences for high instructional spending or because they prefer other characteristics that are correlated with this, perhaps simply academically strong (or high-family-income) peers (Hendricks, Herrington, and Schoellman 2021; Rothstein 2006). Since 1970, deviations from that assortative matching "overwhelmingly result from choices made by students and their families, not choices made by college admissions offices" (Dillon and Smith 2017).

One factor that partially offset the stratifying effect of test-based admissions was the growing concern about diversity. Both public and private universities broadened their scopes in the 1960s and 1970s, with 67 institutions admitting their first female students between 1960 and 1975 (Truffa and Wong 2025) and many institutions admitting nonnegligible numbers of Black undergraduates for the first time. However, schools that wanted to admit more Black students but that relied on standardized tests for admissions faced a problem: Due to the legacy of racial discrimination and segregation, very few Black students exceeded the usual admissions bar at selective colleges. Colleges began to implement affirmative action programs whereby Black students were admitted despite not having the pre-college academic achievement that would have been required of a white applicant. Figure 3 shows that the share of Black students at top-50 private universities more than doubled between 1968 and 1976, with a similar proportional gain (from a lower base) at flagship public universities.[10] Affirmative action also provides somewhat smaller admissions preferences to Hispanic students (Bleemer 2022), and their share at selective colleges also grew rapidly between 1968 and 1976 (though this is somewhat confounded by their growing population share). Figure 3 shows that Asian students were already overrepresented at selective universities in 1968 (when the Asian population share was 0.8 percent) and are substantially overrepresented at those schools today relative to the US population. Affirmative action programs never grew proportionally larger than in this 1970s period—indeed, the Black enrollment share at highly selective private schools hardly changed between 1976 and 2019—but were highly controversial. We discuss such programs further below.

[9] While several studies have emphasized universities' race to enroll high-testing students as being partly motivated by the skewed incentives of public reputational rankings (Espeland and Sauder 2007; Bleemer et al. 2023), this transition notably predates such rankings: *US News & World Report* released its first reputational ranking of universities in 1983.

[10] See Supplemental Appendix Table A-2 for numerical values.

Figure 3
Undergraduate Racial Composition by Tier over Time



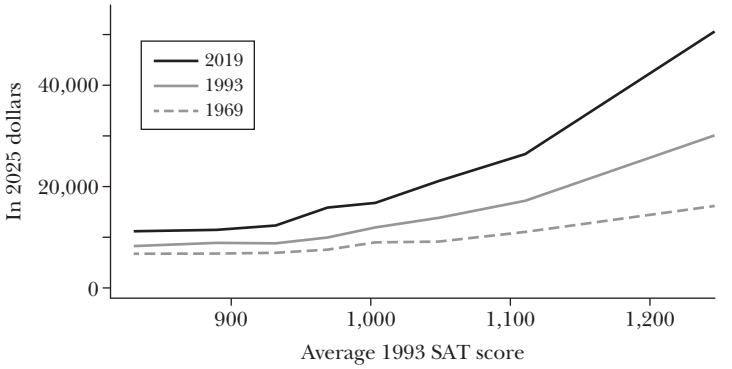
Source: McVay (1969), HEGIS, and IPEDS.
Note: Share of full-time undergraduate students who are Black, Hispanic, or Asian at all four-year universities and by university tier. Supplemental Appendix Tables A-3 and A-4 identify the top-50 private and state flagship universities. "All four-year universities" includes some two-year institutions in 1968 (namely, those that do not have "junior" or "community" in their name), due to data limitations.

Growing Resource Gaps: Since 1970

The cross-university test score distribution had largely stabilized by 1970: For example, the mean test score gap between the bottom and top deciles across universities has remained between 2 and 2.5 national standard deviations (Supplemental Appendix Figure A-7). The share of elite college students from high-income families declined between 1940 and 1980, but this had reverted by 2000 (Abramitzky, Kowalski, Pérez, and Price 2024; Bleemer and Quincy 2025).

However, long-standing funding gaps between universities grew sharply in the last half century. The coefficient of variation (again, the standard deviation divided by the mean) of public universities' per-student instructional expenditures rose from 0.3 in 1969 to over 0.5 in the mid-1990s (as shown earlier in Table 2). Flagship public universities were spending about 35 percent more on instruction than

Figure 4
Instructional Expenditures per Student by Institutions' Average SAT Score in 1993



Source: HEGIS, IPEDS, and the College Board Annual Survey of Colleges.
Note: Average freshman-enrollment-weighted instructional expenditures per full-time equivalent student by eight quantile bins of institutions' 1993 average SAT score, among the 559 four-year institutions with reported 1993 test scores and 1969, 1993, and 2019 enrollments and instructional expenditures. See Supplemental Appendix A for variable definitions.

nonflagships in 1969; by 2008, the gap had grown to 54 percent. As one illustration of what this meant, consider salaries at California's public universities. In 1960, the year of the California Master Plan, the gap in mean wages between full professors in the higher-testing University of California system and the lower-testing California State University system was 23 percent (CCHE 1966). By 2023, the gap had grown to 124 percent, with California State University wages closely tracking inflation and University of California wages growing at a much faster pace (Freedman 2023; University of California 2023).

Figure 4 shows average instructional expenditures per student by institutional average SAT score in 1969, 1993, and 2019. To focus on changes in resources at the institution level, institutions' SAT scores are held fixed here at their 1993 level.[11] In 1969, per-student instructional expenditures at the four-year institutions in the top testing quartile were already just over double those of institutions in the bottom quartile. This national gap widened over time, to almost 150 percent higher in 1993 and 200 percent higher in 2019. Stratification in instructional expenditures has risen more in the private university sector (resulting in a 310 percent expenditure gap between the top and bottom testing terciles in 2019) than the public sector (where the gap is 130 percent), and has been mirrored by rising stratification in both research and administration expenditures. These gaps grew simultaneously with substantial growth in the demand for high-testing institutions (Bound, Hershbein,

[11] The Supplemental Appendix includes a number of figures (A-7 to A-11) that offer robustness checks and additional perspectives on the issues discussed here.

and Long 2009) and in those institutions' observational wage value added (Bleemer and Quincy 2025).

An examination of the revenue changes that financed the growth in the expenditure gap between top- and bottom-test-quartile institutions from 1993 to 2019 reveals evidence of the causes of that growth. The per-student instructional expenditure gap grew by approximately $10,000 over this period. The high-testing institutions saw relative increases in net tuition revenues (by $7,900 per student in the top-quartile relative to bottom-quartile schools), while relative per-student government appropriations actually declined (by $2,100 per student). This was a period of substantial state disinvestment from public higher education in terms of states' block funding allocations (Ehrenberg 2012), and higher-testing universities were relatively more successful at privatizing and raising greater replacement revenue from their students. The remainder of the gap is largely explained by the portions of increased revenue from private gifts and grants (at least $1,200 per student) and government research contracts ($7,800 per student) that subsidize the universities' instructional activities, suggesting that rising federal scientific research investment and philanthropy at high-testing institutions has meaningfully spilled over into those schools' instructional expenditures.

The stratification of instructional resources by standardized test score would look even more extreme if two-year colleges were included in these trends. Community colleges have always had lower average instructional expenditures than their states' nonflagship public universities, by about 40 percent since at least the 1970s, and the wage value added of community colleges is considerably lower (Mountjoy 2022). Two-year college enrollment grew from less than 10 percent of four-year enrollment in 1930—at junior colleges in over 400 cities (Bleemer and Quincy 2026)—to about half of four-year enrollment in the 2010s. A small share of high-demand community college programs implement "meritocratic" admissions policies (Grosz 2023), but their negligible tuition costs and open-access admissions policies generally make them broadly available, which has resulted in increasingly disproportionate enrollment among lower-income and lower-testing students (Kane and Rouse 1999; Bleemer and Quincy 2025).[12]

In sum, these trends reveal that over the past century, American higher education has become dramatically more stratified by student achievement, with increasing relative expenditures at the institutions that high-achieving students attend. This dual trend is what we call the "meritocratic consensus."

[12] Of course, our noting that community colleges have lower value added than four-year colleges should not be taken to suggest that they are not worthwhile for students unable or unprepared to attend a four-year college (Kane and Rouse 1999).

Deviations from "Meritocratic" Admissions

While we have told a story of rising stratification of students by academic achievement and growing funding disparities across colleges and universities, there are some exceptions to these trends. Some highly visible admissions priorities—for diversity (as discussed above), athletes, or children of alumni and donors—cut against the increasing selectivity on student academic backgrounds. Universities often finesse these admission choices by redefining "merit" to include them, but for our purposes they are importantly distinct. However, as we argue below, none of these deviations from academic understandings of merit have been quantitatively large enough to alter our understanding of admissions goals significantly, at least until recently.

Athletics and Other Special Skills

Many selective universities set aside admissions slots for students with special nonacademic skills that are valued for their contributions to campus life. The most prominent such example is collegiate athletics. Athletes typically have poorer pre-college academic achievement than their nonathlete peers at selective universities. For example, the median SAT score of 2014 freshman athletes at the University of California, Berkeley, was at the tenth percentile of the nonathlete SAT distribution (as calculated by the authors; see Bleemer 2018).

At a few small colleges, athletes are a substantial share of the student body: 36 percent at Bowdoin and 47 percent at Claremont McKenna in 2023, for example (US Department of Education, n.d.). However, the athlete share of selective institutional enrollment as a whole is much smaller. Only about 7 percent of college students were intercollegiate athletes in 2023, and only 2 percent were NCAA Division I athletes eligible for preferential tuition pricing and other benefits. Fewer than 10 percent of students at research-intensive R1 public universities or at top-50 private universities are athletes. As a result, athletics contribute very little to the overall allocation of students to US universities.

Some universities also set aside slots for other special talents, such as music. There are many fewer of these slots, and academic standards are not lowered as much (if at all) for them. For example, about 20 freshmen and sophomores (0.3 percent) are members of the University Symphony Orchestra at the University of California, Berkeley, each year, but their test scores were at the seventy-fourth percentile of the nonmusician SAT distribution.

University Community and Finances

Over 80 percent of the top-50 private institutions provide admissions advantages to the relatives of alumni, donors, and faculty (Hurwitz 2011), although these shares have declined in recent years. The justifications offered for these policies are as benefits for institutional stakeholders and incentives for financial support for the institution (Castilla and Poskanzer 2022).

At Harvard in 2014, children of alumni, donors, and faculty were 5.2, 6.9, and 4.1 times more likely, respectively, to be admitted than other academically similar

applicants, respectively (Arcidiacono, Kinsler, and Ransom 2022). Fully 14 percent of admitted students were the children of alumni (Arcidiacono 2018). This vastly overstates the size of the alumni preference, however, as "legacy" applicants also have relatively strong academic credentials. Chetty, Deming, and Friedman (2026) report that, due to preferences for the children of alumni, an unnamed Ivy League institution enrolls about 3 percentage points more students from the top 1 percent of the parental income distribution than it would have if admissions were conducted strictly using standardized test scores.

Fewer than 10 percent of enrollment-weighted research-intensive R1 public universities employ such preferences (Hurwitz 2011). As a result, legacies are relatively limited in their overall impact except at a small number of mostly private institutions.

Affirmative Action

Since the 1960s, many selective universities have provided admissions advantages to students with characteristics like race or class that would otherwise be underrepresented relative to the local or national population. Such policies are most commonly justified by the claim that traditional metrics understate these students' "merit" as they do not take into account the students' poorer pre-college academic opportunity and circumstances (for example, Walton, Spencer, and Erman 2013).

The most prominent example of diversity-based admissions preferences is race-based affirmative action. Standard measures of pre-college academic achievement differ sharply by race (Grodsky, Warren, and Felts 2008). If every Black high school graduate with at least a 1400 SAT score were to attend one of the 50 American colleges and universities where at least 75 percent of enrollees had a 1400 SAT in 2023, those colleges would be just 3 percent Black—less than one-fifth of the 16 percent Black share of high school graduates (Lemann 2024). In fact, 8 percent of those first-year students at those colleges were Black. Among universities that rejected at least 85 (50) percent of their applicants in 2023, 76 (47) percent provided race-based preferences in 2023 (Reber, Goodman, and Nagashima 2023). Preferences had substantial effects on Black and Hispanic students' admissions chances: One study found that in the 1990s, applicants from underrepresented minorities to the University of California campuses in Berkeley and Los Angeles were twice as likely to be admitted as other academically comparable applicants (Bleemer 2023).

Ten states (including California) banned race-based preferences between 1997 and 2020, and the Supreme Court banned them nationwide in 2023 (*Students for Fair Admissions v. Harvard*, 600 US 181 [2023]). As a result, colleges no longer consider race explicitly in admissions (though they may consider other factors that correlate with it).

Some universities have also used affirmative action to manage their gender distributions. Female academic performance generally outpaces that of males in high school (Reeves 2022). As a result, gender-blind admissions would typically yield a class that is much more than 50 percent female at many highly selective colleges.

Baum and Goodstein (2005) find evidence of male admissions preferences in the late 1990s at small liberal arts colleges (though see also Bielby, Posselt, Jaquette, and Bastedo 2014). The impact of male affirmative action is quantitatively too small to play a significant role in the aggregate allocation of students to universities (see Supplemental Appendix B).

Holistic Review and Alternative Metrics

Some less-selective colleges use formulas or cutoffs for admissions (Zimmerman 2014; Mountjoy 2025; Miller 2025), but nearly all highly selective public and private universities conduct admissions by "holistic review" (Bastedo, Bowman, Glasener, and Kelly 2018)—that is, rather than admitting students using a fixed-weight ranking of students by academic criteria (such as a minimum SAT threshold or SAT-GPA combination), admissions decisions are made by a "trained evaluator or set of evaluators [who] craft a single score for the applicant based upon a combination of [academic and nonacademic] criteria" (BOARS 2012).

Holistic review reallocates admissions weight from test scores to essays and other components. The latent weights that admissions readers place on the different application components in holistic review likely vary across institutions, but in most cases it appears that scoring remains based primarily on an assessment of academic achievement (Mare 2014; Arcidiacono 2018). Zitzewitz (2025) shows that switching to holistic review can affect the admissions decisions of 10–15 percent of admitted students and result in a decline in average test scores. However, this probably overstates the effect on selection by academic achievement, as in many cases holistic review merely replaces students with high standardized test scores with others whose applications show other academic achievement indicators (like strong high school grades).

Five states' public university systems combine test-based or holistic review with "top percent" policies, which provide an admissions advantage to the highest-ranking (by high school grade point average) graduates from every eligible high school, regardless of their test scores: Texas, California, Florida, Georgia, and New York.[13] The net effect of these policies is to provide admissions advantages to high-achieving students from low-performing high schools, because top students at high-performing schools generally have high test scores and would be admitted under traditional rules (Black, Denning, and Rothstein 2023). They have been found to slightly increase enrollment of underrepresented minorities (by 1–2 percentage points; Kapor 2025; Bleemer 2021) and enrollment from families with below-median income at participating selective universities (as much as 1 percentage point; Bleemer 2023).

Admissions to selective US colleges clearly involve more than standardized test scores. Some students are admitted for their athletic ability or personal connections, while others get boosts for increasing diversity. But most college and university

[13] These policies have a precedent in the "highest seventh plan" implemented by Harvard and other private Massachusetts colleges in the 1920s (Hinckley 1941).

admissions decisions, while not formulaic, remain based on demonstrated academic achievement.

Is Merit-Based Stratification Efficient?

The American higher education system provides substantially greater educational resources to students with relatively stronger pre-college academic achievement. Is this allocation efficient, in the sense of allocating students and resources to schools in a way that maximizes some understanding of social welfare? This cannot be assumed; the higher education sector is dominated by imperfect information, reputation effects, nonprice rationing, and nonprofit and state actors, and no welfare theorem ensures that the market outcome is efficient. There might well be other allocations of students to schools, or of resources to students, that would achieve higher social welfare, and perhaps even that could be sustained as equilibria given the behavior of students and institutions.

Higher education can be seen as a production process where institution-provided inputs like classrooms, dormitories, and instructors are combined with students to produce learning output. We set aside the question of the relevant dimensions of output—is the goal to maximize earnings, or are there other forms of learning that should be pursued?—and assume an agreed-upon, unidimensional output. Each university produces this output by combining institutional inputs with students, who we assume can also be summarized by a single index: their average pre-college academic achievement (setting aside questions considered above about how to measure this). The production function can be quite general, allowing for various interactions. In particular, the dependence of output on student characteristics allows for a range of types of peer effects, where the outcomes of one student depend on the characteristics of classmates.

One initial question is whether it is efficient for students to be so stratified at all by pre-college academic achievement. In many countries (like Canada and Germany), there is much less stratification, with students generally attending their local college. There is little evidence on the efficiency of "tracking" in higher education; the large literature on K-12 tracking suggests that increasing the variance of student achievement increases inequality but has little effect on average outcomes (Betts 2011; Terrin and Triventi 2023).

We focus on a second question: Given substantial stratification, is it efficient for resources to be allocated preferentially to the highest-achieving students? We consider three theories about why it might be.

q-Complementarity

Two inputs in production are *q*-complements, where *q* refers to "quantity," if increasing the amount of one raises the marginal product of the other. The most straightforward rationalization of the meritocratic consensus would be if pre-college academic achievement and per-student instructional spending are *q*-complements

in the production of human capital. This model has some plausibility: Perhaps higher-achieving students benefit more from small classes (which facilitate individualized exploration), professors at the research frontier who can teach more difficult course material, and more up-to-date lab facilities. If so, social gains can arise from segregating those high-achieving students from others and directing substantial resources in their direction (Sallee, Resch, and Courant 2008). This creates an equity-efficiency tradeoff: It is efficient to direct more resources to the highest-achieving students, but because high-achievement students tend to come from advantaged families, such an allocation is regressive.

Benabou (2000) notes that in the presence of *q*-complementarity, the meritocratic consensus may, in effect, be antimeritocratic if meritocracy is understood as maximizing the share of income variation arising from variation in innate ability (as opposed to family background). A traditionally meritocratic allocation would direct higher instructional spending to the schools that high-ability disadvantaged students attend, which might be schools with low average pre-college academic achievement if that is sufficiently correlated with family background (see related discussion in Roemer 1998).

An alternative view of the higher education production process, in which pre-college educational attainment and per-student spending are *q-substitutes* rather than *q*-complements, is also plausible. Perhaps *weaker* students benefit more from individualized attention, out-of-classroom advising, tutoring, and student support services, while stronger students can learn nearly as effectively with only a lecture and a textbook. Elementary and secondary education funding policies are often based on an assumption like this about the production process; for example, "weighted student funding" formulas direct more resources to schools serving disadvantaged students (for example, Chingos and Blagg 2017).

If student aptitude or preparedness is a *q*-substitute for school resources, it might still make sense to segregate stronger from weaker students (depending on the nature of peer effects), but the efficient allocation would direct *more* resources to the schools where weaker students enroll. In public policy terms, such a policy might involve more generous funding for community colleges and open access institutions, and lower funding—at least for instruction—at state flagships. However, there is little sign that policymakers prefer such funding arrangements.

Convex Social Welfare

A second model in which the observed allocations might be efficient is one in which the social welfare function puts very high weight on the human capital of the highest achievers—in technical terms, if the social welfare function is convex in students' post-college human capital. A possible justification for such a model would be if these top achievers are responsible for a disproportionate share of scientific breakthroughs, entrepreneurial successes, and artistic masterpieces. In this case, it could then be optimal to concentrate the best students with the highest educational resources, not because they receive the highest human capital benefit from these

resources, but because any human capital increment that they receive is disproportionately valuable to society.

High School Incentivization

A third model emphasizes educational production in high school rather than in college. Students decide how much effort to exert on their high school studies, and policies that increase the reward for high achievement can increase student effort (Coate and Loury 1993; Fryer, Loury, and Yuret 2008; Akhtari, Bau, and Laliberté 2024). Collegiate resources can be such a reward, even if they are substitutes for student preparedness in the *college* production process. This follows the same logic as meritocracy in competitive labor markets: When firms hire and promote the most productive workers, the resulting wage dispersion encourages effort and allocative efficiency (for example, Holmström 1979). The incentive effects of large rewards for high school achievement could theoretically induce sufficient effort in high school to justify even inefficient allocations in college.

Evidence on the Theories

What does the empirical evidence tell us about these theories? There is a fair amount of available evidence regarding the *q*-complementarity theory. Many scholars have studied whether the wage return to *more-selective* university enrollment is higher for higher-testing students, focusing in particular on whether some lower-testing students would be "mismatched" with the demanding curriculum of more-selective universities and, thus, better off enrolling elsewhere. For example, an early series of observational and structural studies of race-based affirmative action—which admits lower-testing underrepresented minority students to more-selective universities—provided mixed evidence that the policy mismatched students (for example, Loury and Garman 1995; Arcidiacono and Lovenheim 2016; Rothstein and Yoon 2008). A more recent quasi-experimental literature has found that, if anything, the benefits of university selectivity *decrease* in pre-college academic achievement for reasonably prepared students (Bleemer 2022; Barahona, Dobbin, and Otero 2023). This finding is mirrored in broader studies of the return to university selectivity: Lower-testing students generally benefit as much as (Mountjoy and Hickman 2021; Chetty, Deming, and Friedman 2026) or more than (Dale and Krueger 2014; Bleemer 2021) higher-testing students from attending more selective colleges (though, see Dillon and Smith 2020 for a contrary conclusion). However, these studies, whatever their findings, generally do not distinguish between peer groups and resources as determinants of college value added.[14]

[14] Models of dynamic complementarity ("skill begets skill") in K-12 educational production (Cunha and Heckman 2007; Johnson and Jackson 2019) support the view that resources are complementary with prior achievement. However, several studies have found that small classes are more beneficial for lower-achieving students (Krueger 1999; Summers and Wolfe 1977), where dynamic complementarity would imply the opposite. In any case, the postsecondary production function may differ from that of K-12.

Other evidence comes from specific program evaluations. Several resource-intensive programs have yielded large benefits for college students with low prior achievement. An initiative called the Accelerated Study in Academic Programs provides financial, academic, and personal support to students at the City University of New York and significantly raises student success (Scrivener et al. 2015). Similar evidence comes from the Opening Doors demonstration project, aimed at improving higher education outcomes for community college students (Scrivener et al. 2008; Scrivener and Coghlan 2011). However, it is difficult to compare the benefits of these programs to the potential benefits of devoting similar resources to high-prior-achievement students. Similarly, research on "massively open online courses," which have attracted attention as a potentially less resource-intensive way to deliver postsecondary instruction, has found that this model was less effective for students with low prior achievement—which again suggests that resources and preparedness may be *q*-substitutes rather than complements (Hansen and Reich 2015; Kofoed, Gebhart, Gilmore, and Moschitto 2024).

Turning to the other two models, there is very little evidence on the convexity of the relationship between human capital and social value. There is a positive convex relationship between childhood test scores and innovation, as measured by patenting (Bell et al. 2019), but this could be a direct effect of individual characteristics or family background rather than a consequence of school-constructed human capital. The relationship between entrepreneurial success and SAT scores, on the other hand, is positive but slightly *concave* (Chetty et al. 2025), even in an environment of highly stratified college resources. This suggests that high-testing students are unlikely to become disproportionately more likely to found successful firms as a result of enrolling at high-resource institutions. Finally, there is little available evidence on the potential magnitude of incentive effects of college quality on high school effort or achievement.

Considering these three possible justifications for the meritocratic consensus as a whole, there is no compelling evidence that it is efficient to direct the most resources to the most highly prepared students. What evidence there is offers weak support for the view that resources can substitute for preparedness and suggests that, at current margins, overall learning would be increased by shifting resources to schools that enroll students with lower prior achievement.

The Future of the Meritocratic Consensus

Several recent developments seem likely to affect the scope of the meritocratic consensus in the short and medium term.

Funding Allocation

The shift toward higher funding at more selective institutions has leveled out or even become less pronounced since about 2010 (Supplemental Appendix Figure A-12). The between-institution associations between average SAT score and

both net tuition revenue and private funding have continued their slow growth in recent years, in part likely driven by rising international undergraduate and graduate enrollment at more-selective institutions (Bound, Braga, Khanna, and Turner 2021). However, the relationship between average SAT score and government appropriations has become meaningfully negative over the past two decades. This has been almost wholly driven by increasing funding from student grant programs to low-SAT institutions, not by changes in states' block appropriations. State financial aid programs like New York's Excelsior Scholarship and New Jersey's Tuition Aid Grant have played an important role in this nascent trend.[15] Recent federal actions targeting the finances of high-revenue private universities, such as the expanded university endowment tax in the One Big Beautiful Bill Act of 2025, could compound this trend, though whether it continues will depend on federal and state legislative priorities.

College Admissions

The early 2020s brought two of the most significant changes to university admissions policies in over 50 years. First, after about 25 percent of private universities adopted "test-optional" admissions policies between the 1990s and 2010s, the 2020 COVID-19 pandemic then forced many other institutions to drop their test requirements. The share of "test-mandatory" institutions fell to just 8 percent in 2023, comparable to its level in the 1930s. Some universities went test-optional, meaning that applicants could decide whether to submit standardized test scores, while others (including all public universities in California) have gone test-blind, prohibiting applicants from submitting scores. The Ivy League schools largely reinstated their testing requirements between 2024 and 2026, but as of mid-2026, most top-50 private and flagship public institutions had not.

The effect of test-optional and test-blind policies on student stratification by pre-college academic achievement is unknown—not least because tests are the most common measure of cross-university differences in academic achievement—but their widespread adoption may herald meaningful changes in the centrality of academic achievement in university admissions. Schools that implemented test-optional admissions prior to 2016 saw the stabilization of previously declining application rates and a 1–2 percentage point increase in their enrollment shares of underrepresented minorities (Bennett 2022).

Second, the US Supreme Court decision in *Students for Fair Admissions v. Harvard* (600 US 181 [2023]) prohibited race-based affirmative action in college admissions. The year following the Court decision, first-year enrollment of underrepresented minorities at top-50 private universities declined from 22 to 18 percent. However, an enrollment cascade of students into less-selective institutions (Conrad and Sharpe 1996) led 2024 enrollment of first-year underrepresented minority students at flagship public universities to rise by 3 percent. The downstream effects

[15] The Supplemental Appendix includes a number of figures (A-12 to A-14) that offer robustness checks and additional perspectives on the issues discussed here.

of the *SFFA* decision—and related actions by the Trump administration to obligate greater "meritocratic" selection in university admissions—are still evolving.

Within-Institution Stratification

While enrollment stratification by levels of pre-college academic achievement has been largely stable for decades and inequity in resource allocations between institutions with higher and lower test scores has stopped growing, recent decades have seen the increasing allocation of greater resources to higher-achieving students *within* institutions.

As one example, college students in the early 1960s at nearly all institutions could declare any available major in which they could earn passing grades. In recent decades, however, many institutions (and especially public universities) have begun rationing access to high-demand majors, often using admissions processes that resemble those used for college admissions. Among lucrative majors like engineering and business, over 25 percent of enrollment-weighted programs at the "R1" research-intensive public universities have minimum grade-point-average requirements in introductory courses of above 2.3 (higher than the passing grade of 2.0) and another 20 percent implement admissions policies mirroring holistic review (based on grades, essays, and other materials). Between 2010 and 2024, the share of computer science programs at R1 public universities—often the institution's highest-value major—that restricted access on the basis of academic preparation rose from 10 to 40 percent (Bleemer and Mehta 2024). "Soft" restrictions like relatively low grade distributions in courses related to science, technology, engineering, and mathematics have also become more prevalent (for example, Butcher, McEwan, and Weerapanda 2014).

These policies have contributed to substantial between-major stratification by academic achievement, with higher-achieving students allocated to higher-value majors. This in turn manifests as swiftly growing gaps between the average relative value of college majors earned by more- or less-advantaged students (Bleemer and Mehta 2024; Bleemer and Quincy 2025). As with allocations across institutions, there is little evidence that within-institution "meritocratic" allocation of majors is efficient, with the available evidence instead suggesting that providing lucrative major access to *lower*-achievement students would result in greater human capital production (Bleemer and Mehta 2022).

The growing prevalence of "honors colleges"—separate programs within large universities that provide additional resources to selected high-achieving students that are not available to the broader student body—is another recent avenue by which resources are preferentially allotted to higher-performing students within institutions. Between 1958 and 2021, the number of research-intensive universities with honors colleges rose from three to about 200, with growth accelerating over time. Honors colleges generally provide dedicated curricula in small classrooms, special advising services, and additional student amenities, typically without any additional cost for the student (Cognard-Black and Smith 2023). While there is little evidence on the long-run effects of honors college participation, studies of "gifted

education" programs at the K-12 level suggest that expanding program access to lower-performing students would provide those students substantial long-run value (Card, Chyn, and Giuliano 2024; Goldstein, Lavy, and Rachkovski 2026).

Conclusion

Colleges and universities in the United States are highly stratified by students' pre-college academic achievement, with the highest-achieving students attending colleges that spend dramatically more on instruction than those that lower-achieving students attend. Neither of these is a necessary feature of a higher education market; universities need not be as stratified by academic achievement as they are in the United States, nor do resources need to be so inequitably distributed. There is little economic evidence justifying the status quo on efficiency grounds, and substantial room for skepticism. Whether this feature of the American higher education system persists depends on whether universities—and their primary owners, the American public—wish it to.

■ *We thank Sarah Turner for helpful comments and Dylan Black, Ivy Chen, and Shinrea Su for excellent research assistance.*

References

Abramitzky, Ran, Jennifer K. Kowalski, Santiago Pérez, and Joseph Price. 2024. "The GI Bill, Standardized Testing, and Socioeconomic Origins of the US Educational Elite over a Century." NBER Working Paper 33164.

Akhtari, Mitra, Natalie Bau, and Jean-William Laliberté. 2024. "Affirmative Action and Pre-college Human Capital." *American Economic Journal: Applied Economics* 16 (1): 1–32.

Arcidiacono, Peter. 2018. *Expert Report of Peter S. Arcidiacono, Students for Fair Admissions, Inc. v. Harvard* No. 14-cv-14176-ADB(D. Mass.).

Arcidiacono, Peter, Josh Kinsler, and Tyler Ransom. 2022. "Legacy and Athlete Preferences at Harvard." *Journal of Labor Economics* 40 (1): 133–56.

Arcidiacono, Peter, and Michael Lovenheim. 2016. "Affirmative Action and the Quality–Fit Trade-Off." *Journal of Economic Literature* 54 (1): 3–51.

Barahona, Nano, Cauê Dobbin, and Sebastián Otero. 2023. "Affirmative Action in Centralized College Admissions Systems." Unpublished.

Bastedo, Michael N., Nicholas A. Bowman, Kristen M. Glasener, and Jandi L. Kelly. 2018. "What Are We Talking about When We Talk about Holistic Review? Selective College Admissions and Its Effects on Low-SES Students." *Journal of Higher Education* 89 (5): 782–805.

Baum, Sandy, and Eban Goodstein. 2005. "Gender Imbalance in College Applications: Does It Lead to a Preference for Men in the Admissions Process?" *Economics of Education Review* 24 (6): 665–75.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. 2019. "Who Becomes an Inventor in America? The Importance of Exposure to Innovation." *Quarterly Journal of Economics* 134 (2): 647–713.

Bénabou, Roland. 2000. "Meritocracy, Redistribution, and the Size of the Pie." In *Meritocracy and Economic Inequality*, edited by Kenneth Arrow, Samuel Bowles, and Steven Durlauf, 317–39. Princeton University Press.

Bender, William J. 1952. "Admission Policy Memo." In *Admissions 1957–1958: Report*, General Harvard Admissions Collection (HUC 957.3), Harvard University Archives, Cambridge, MA.

Bennett, Christopher T. 2022. "Untested Admissions: Examining Changes in Application Behaviors and Student Demographics under Test-Optional Policies." *American Educational Research Journal* 59 (1): 180–216.

Betts, Julian R. 2011. "The Economics of Tracking in Education." In *Handbook of the Economics of Education*, Vol. 3, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 341–81. North-Holland.

Bielby, Rob, Julie Renee Posselt, Ozan Jaquette, and Michael N. Bastedo. 2014. "Why Are Women Underrepresented in Elite Colleges and Universities? A Non-linear Decomposition Analysis." *Research in Higher Education* 55 (8): 735–60.

Black, Sandra E., Jeffrey T. Denning, and Jesse Rothstein. 2023. "Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes." *American Economic Journal: Applied Economics* 15 (1): 26–67.

Bleemer, Zachary. 2018. "The UC ClioMetric History Project and Formatted Optical Character Recognition." Center for Studies in Higher Education Research Paper 3.18.

Bleemer, Zachary. 2021. "Top Percent Policies and the Return to Postsecondary Selectivity." Center for Studies in Higher Education Research Paper 1.21.

Bleemer, Zachary. 2022. "Affirmative Action, Mismatch, and Economic Mobility after California's Proposition 209." *Quarterly Journal of Economics* 137 (1): 115–60.

Bleemer, Zachary. 2023. "Affirmative Action and Its Race-Neutral Alternatives." *Journal of Public Economics* 220: 104839.

Bleemer, Zachary, and Jesse Rothstein. 2026. *Data and Code for: "The Meritocratic Consensus and Stratification in Higher Education."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. Https://doi.org/10.3886/E259616V1.

Bleemer, Zachary, Mukul Kumar, Aashish Mehta, Chris Muellerleile, and Christopher Newfield. 2023. *Metrics That Matter: Counting What's Really Important to College Students*. Johns Hopkins University Press.

Bleemer, Zachary, and Aashish Mehta. 2022. "Will Economics Make You Rich? A Regression Discontinuity Analysis of the Return to College Major." *American Economic Journal: Applied Economics* 14 (2): 1–22.

Bleemer, Zachary, and Aashish Mehta. 2024. "College Major Restrictions and Student Stratification." NBER Working Paper 33269.

Bleemer, Zachary, and Sarah Quincy. 2025. "Changes in the College Mobility Pipeline Since 1900." NBER Working Paper 33797.

Bleemer, Zachary, and Sarah Quincy. 2026. "Junior Colleges and Local Development." Unpublished.

BOARS. 2012. *Comprehensive Review in Admissions at the University of California: An Update*. University of California. https://senate.universityofcalifornia.edu/_files/committees/boars/BOARSREPORTCOMPREHENSIVEREVIEW2012.pdf.

Bound, John, Breno Braga, Gaurav Khanna, and Sarah Turner. 2021. "The Globalization of Postsecondary Education: The Role of International Students in the US Higher Education System." *Journal of Economic Perspectives* 35 (1): 163–84.

Bound, John, Brad Hershbein, and Bridget Terry Long. 2009. "Playing the Admissions Game: Student Reactions to Increasing College Competition." *Journal of Economic Perspectives* 23 (4): 119–46.

Bound, John, Michael Lovenheim, and Sarah Turner. 2010. "Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources." *American Economic Journal: Applied Economics* 2 (3): 129–57.

Bound, John, and Sarah Turner. 2007. "Cohort Crowding: How Resources Affect Collegiate Attainment." *Journal of Public Economics* 91 (5–6): 877–99.

Butcher, Kristin F., Patrick J. McEwan, and Akila Weerapanda. 2014. "The Effects of an Anti-Grade-Inflation Policy at Wellesley College." *Journal of Economic Perspectives* 28 (3): 189–204.

Card, David, Eric Chyn, and Laura Giuliano. 2024. "Can Gifted Education Help Higher-Ability Boys from Disadvantaged Backgrounds?" NBER Working Paper 33282.

Carnegie Foundation for the Advancement of Teaching. 2000. *The Carnegie Classification of Institutions of Higher Education, 2000 Edition*. Carnegie Foundation for the Advancement of Teaching.

Castilla, Emilio J., and Ethan J. Poskanzer. 2022. "Through the Front Door: Why Do Organizations (Still) Prefer Legacy Applicants?" *American Sociological Review* 87 (5): 782–826.

Coordinating Council for Higher Education. 1966. *Annual Report on Faculty Salaries, Benefits, and Recruitment*. Coordinating Council for Higher Education.

Chetty, Raj, David J. Deming, and John N. Friedman. 2026. "Diversifying Society's Leaders? The Determinants and Causal Effects of Admission to Highly Selective Private Colleges." *Quarterly Journal of Economics* 141 (1): 51–145.

Chetty, Raj, Gaia Dossi, Matt Smith, John van Reenen, Owen Zidar, and Erick Zwick. 2025. "Creating New Businesses in America: The Determinants of and Returns to Entrepreneurship." Unpublished.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2020. "Income Segregation and Intergenerational Mobility across Colleges in the United States." *Quarterly Journal of Economics* 135 (3): 1567–633.

Chingos, Matthew M., and Kristin Blagg. 2017. *Do Poor Kids Get Their Fair Share of School Funding?* Urban Institute.

Clothier, Robert. 1946. "Letter to Hon. Mattie S. Doremus." In Records of the Robert C. Clothier Administration, 1925–1952 (RG 04/A16), Box 1, Folder 3, Rutgers University Archives, New Jersey.

Coate, Stephen, and Glenn C. Loury. 1993. "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *American Economic Review* 83 (5): 1220–40.

Cognard-Black, Andrew J., and Patricia Joanne Smith. 2023. "Characteristics of the 21st-Century Honors College." In *Honors Colleges in the 21st Century*, edited by Richard Badenhausen, 23–79. National Collegiate Honors Council.

Cohodes, Sarah R., and Joshua S. Goodman. 2014. "Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy." *American Economic Journal: Applied Economics* 6 (4): 251–85.

Committee of Ten. 1892. *Report of the Committee of Ten on Secondary School Studies*. American Book Company on behalf of the National Educational Association.

Conrad, Cecilia A., and Rhonda V. Sharpe. 1996. "The Impact of the California Civil Rights Initiative (CCRI) on University and Professional School Admissions and the Implications for the California Economy." *Review of Black Political Economy* 25 (1): 13–59.

Coons, Arthur G., Arthur D. Browne, Howard A. Campion, Glenn S. Dumke, Thomas C. Holy, Dean E. McHenry, Henry. T. Tyler, and Robert J. Wert. 1960. *A Master Plan for Higher Education in California, 1960–1975*. California State Department of Education.

Cooper, William John. 1931. *Biennial Survey of Education, 1928–1930, Volume II*. US Government Printing Office.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47.

Dale, Stacy Berg, and Alan B. Krueger. 2002. "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *Quarterly Journal of Economics* 117 (4): 1491–527.

Dale, Stacy, B. and Alan B. Krueger. 2014. "Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data." *Journal of Human Resources* 49 (2): 323–58.

de Brey, Cristobal, Thomas D. Snyder, Anlan Zhang, and Sally A. Dillow. 2021. *Digest of Education Statistics 2019 (NCES 2021-009)*. National Center for Education Statistics, Institute of Education Sciences, US Department of Education.

Deming, David J., and Christopher R. Walters. 2017. "The Impact of Price Caps and Spending Cuts on U.S. Postsecondary Attainment." NBER Working Papers 23736.

Dillon, Eleanor Wiske, and Jeffrey Andrew Smith. 2017. "Determinants of the Match between Student Ability and College Quality." *Journal of Labor Economics* 35 (1): 45–66.

Dillon, Eleanor Wiske, and Jeffrey Andrew Smith. 2020. "The Consequences of Academic Match between Students and Colleges." *Journal of Human Resources* 55 (3): 767–808.

Ehrenberg, Ronald G. 2012. "American Higher Education in Transition." *Journal of Economic Perspectives* 26 (1): 193–216.

Espeland, Wendy Nelson, and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113 (1): 1–40.

Freedman, Leora. 2023. *The Employees of the California State University, 2023*. California State University Systemwide Human Resources. https://www.calstate.edu/csu-system/faculty-staff/employee-profile/Documents/2023-Employee-Profile-Brochure.pdf.

Fryer, Roland G., Jr., Glenn C. Loury, and Tolga Yuret. 2008. "An Economic Analysis of Color-Blind Affirmative Action." *Journal of Law, Economics, and Organization* 24 (2): 319–55.

Goldin, Claudia, and Lawrence F. Katz. 1999. "The Shaping of Higher Education: The Formative Years in the United States, 1890 to 1940." *Journal of Economic Perspectives* 13 (1): 37–62.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives* 20 (4): 133–56.

Goldstein, Yoav, Victor Lavy, and Genia Rachkovski. 2026. "Gifted Children Programs' Short and Long-Term Impact: Higher Education, Earnings, and the Knowledge Economy." NBER Working Paper 29779.

Grodsky, Eric, John Robert Warren, and Erika Felts. 2008. "Testing and Social Stratification in American Education." *Annual Review of Sociology* 34: 385–404.

Grosz, Michel. 2023. "Admissions Policies, Cohort Composition, and Academic Success: Evidence from California." *Journal of Human Resources* 58 (4): 1242–72.

Hansen, John D., and Justin Reich. 2015. "Democratizing Education? Examining Access and Usage Patterns in Massive Open Online Courses." *Science* 350 (6265): 1245–48.

Hendricks, Lutz, Christopher Herrington, and Todd Schoellman. 2021. "College Quality and Attendance Patterns: A Long-Run View." *American Economic Journal: Macroeconomics* 13 (1): 184–215.

Hinckley, William W. 1941. *Handbook of College Entrance Requirements*. Vol. 13. US Department of Education.

Hoekstra, Mark. 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *Review of Economics and Statistics* 91 (4): 717–24.

Holmström, Bengt. 1979. "Moral Hazard and Observability." *Bell Journal of Economics* 10 (1): 74–91.

Hoxby, Caroline M. 2009. "The Changing Selectivity of American Colleges." *Journal of Economic Perspectives* 23 (4): 95–118.

Hurt, Huber William, ed. 1923–1962. *The College Blue Book*. 1st, 2nd, 3rd, 4th, 5th, and 6th eds. Christian E. Burckel and Associates.

Hurwitz, Michael. 2011. "The Impact of Legacy Status on Undergraduate Admissions at Elite Colleges and Universities." *Economics of Education Review* 30 (3): 480–92.

Jackson, C. Kirabo, and Claire L. Mackevicius. 2024. "What Impacts Can We Expect from School Spending Policy? Evidence from Evaluations in the United States." *American Economic Journal: Applied Economics* 16 (1): 412–46.

Johnson, Rucker C., and C. Kirabo Jackson. 2019. "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending." *American Economic Journal: Economic Policy* 11 (4): 310–49.

Kane, Thomas J., Peter R. Orszag, and Emil Apostolov. 2005. "Higher Education Appropriations and Public Universities: Role of Medicaid and the Business Cycle." In *Brookings-Wharton Papers on Urban Affairs*, edited by Gary T. Burtless and Janet Rothenberg Pack, 99–146. Brookings Institution Press.

Kane, Thomas J., and Cecilia Elena Rouse. 1999. "The Community College: Educating Students at the Margin between College and Work." *Journal of Economic Perspectives* 13 (1): 63–84.

Kapor, Adam. 2025. "Transparency and Percent Plans." *Econometrica* 93 (6): 2123–57.

Karabel, Jerome. 1989. *Freshman Admissions at Berkeley: A Policy for the 1990s and Beyond*. Committee on Admissions and Enrollment at the University of California, Berkeley.

Karabel, Jerome. 2005. *The Chosen: The Hidden History of Admission and Exclusion at Harvard, Yale, and Princeton*. Mariner Books.

Kofoed, Michael S., Lucas Gebhart, Dallas Gilmore, and Ryan Moschitto. 2024. "Zooming to Class? Experimental Evidence on College Students' Online Learning during COVID-19." *American Economic Review: Insights* 6 (3): 324–40.

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114 (2): 497–532.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach. 2018. "School Finance Reform and the Distribution of Student Achievement." *American Economic Journal: Applied Economics* 10 (2): 1–26.

Lemann, Nicholas. 1999. *The Big Test: The Secret History of the American Meritocracy*. Farrar, Straus and Giroux.

Lemann, Nicholas. 2024. *Higher Admissions: The Rise, Decline, and Return of Standardized Testing*. Princeton University Press.

Loury, Linda Datcher, and David Garman. 1995. "College Selectivity and Earnings." *Journal of Labor Economics* 13 (2): 289–308.

Mare, Robert D. 2014. *Holistic Review in Freshman Admissions at the University of California-Los Angeles, 2009–11 Update*. Committee on Undergraduate Admissions and Relations with Schools, University of California, Los Angeles. https://sard.law/static/sard/pdf/Mare%20Report%202014.pdf.

McVay, Ruth. 1969. *Undergraduate Enrollment by Ethnic Group in Federally Funded Institutions of Higher Education*. US Department of Health, Education, and Welfare Office of Civil Rights. https://catalog.hathitrust.org/Record/012336075.

Meder, Albert, Jr. 1939. "Letter to Robert C. Clothier." In Records of the Robert C. Clothier Administration, 1925–1952 (RG 04/A16), Box 77, Folder 15, Rutgers University Archives, New Jersey.

Miller, Lois. 2025. "Switching Schools: Effects of College Transfers." Annenberg Institute at Brown University EdWorkingPaper 25-1159.

Mountjoy, Jack. 2022. "Community Colleges and Upward Mobility." *American Economic Review* 112 (8): 2580–630.

Mountjoy, Jack. 2025. "Marginal Returns to Public Universities." NBER Working Paper 32296.

Mountjoy, Jack, and Brent R. Hickman. 2021. "The Returns to College(s): Relative Value-Added and Match Effects in Higher Education." NBER Working Paper 29276.

Reber, Sarah, Gabriela Goodman, and Rina Nagashima. 2023. "Admissions at Most Colleges Will Be Unaffected by Supreme Court Ruling on Affirmative Action." Brookings Institution, November 7. https://www.brookings.edu/articles/admissions-at-most-colleges-will-be-unaffected-by-supreme-court-ruling-on-affirmative-action/.

Reeves, Richard V. 2022. *Of Boys and Men: Why the Modern Male Is Struggling, Why It Matters, and What to Do about It*. Brookings Institution Press.

Roemer, John E. 1998. *Equality of Opportunity*. Harvard University Press.

Rothstein, Jesse M. 2006. "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions." *American Economic Review* 96 (4): 1333–50.

Rothstein, Jesse. Forthcoming. "The Big Sort: How Privilege Becomes Permanent in the New American Economy. *University of Chicago Press*.

Rothstein, Jesse, and Diane Whitmore Schanzenbach. 2022. "Does Money Still Matter? Attainment and Earnings Effects of Post-1990 School Finance Reforms." *Journal of Labor Economics* 40 (S1): S141–78.

Rothstein, Jesse, and Albert Yoon. 2008. "Mismatch in Law School." NBER Working Paper 14275.

Sallee, James M., Alexandra M. Resch, and Paul N. Courant. 2008. "On the Optimal Allocation of Students and Resources in a System of Higher Education." *B.E. Journal of Economic Analysis and Policy* 8 (1): 11.

Scrivener, Susan, Dan Bloom, Allen LeBlanc, Christina Paxson, Cecilia Elena Rouse, and Colleen Sommo. 2008. *Opening Doors. A Good Start: Two-Year Effects of a Freshmen Learning Community Program at Kingsborough Community College*. MDRC.

Scrivener, Susan, and Erin Coghlan. 2011. *Opening Doors to Student Success: A Synthesis of Findings from an Evaluation at Six Community Colleges*. MDRC.

Scrivener, Susan, Michael J. Weiss, Alyssa Ratledge, Timothy Rudd, Colleen Sommo, and Hannah Fresques. 2015. *Doubling Graduation Rates: Three-Year Effects of CUNY's Accelerated Study in Associate Programs (ASAP) for Developmental Education Students*. MDRC.

Spindt, Herman, Ralph Prator, J. Edward Sander, Basil Peterson, and Grant Jensen. 1961. *Selection and Retention of Students in California's Institutions of Higher Education*. California State Department of Education. https://books.google.com/books?id=TCM8AAAAIAAJ.

Stanley, Marcus. 2003. "College Education and the Midcentury GI Bills." *Quarterly Journal of Economics* 118 (2): 671–708.

Summers, Anita A., and Barbara L. Wolfe. 1977. "Do Schools Make a Difference?" *American Economic Review* 67 (4): 639–52.

Terrin, Éder, and Moris Triventi. 2023. "The Effect of School Tracking on Student Achievement and Inequality: A Meta-Analysis." *Review of Educational Research* 93 (2): 236–74.

Thelin, John R. 2010. "The Attrition Tradition in American Higher Education: Connecting Past and Present." AEI Future of American Education Project Working Paper 2010-01.

Truffa, Francesca, and Ashley Wong. 2025. "Undergraduate Gender Diversity and the Direction of Scientific Research." *American Economic Review* 115 (7): 2414–48.

University of California. 2023. *Compensation at the University of California*. Regents of the University of California. https://ucannualwage.ucop.edu/wage/.

US Bureau of Labor Statistics. 2026. "Consumer Price Index for All Urban Consumers (CPI-U): US City Average, All Items, Annual Averages 1913–2025." https://data.bls.gov/ (accessed July 1, 2026).

US Department of Education. n.d. "Equity in Athletics Data Analysis." https://ope.ed.gov/athletics/#/.

Veysey, Laurence R. 1965. *The Emergence of the American University*. University of Chicago Press.

Walton, Gregory M., Steven J. Spencer, and Sam Erman. 2013. "Affirmative Meritocracy." *Social Issues and Policy Review* 7 (1): 1–35.

Westrick, Paul A., Jessica P. Marini, Linda Young, Helen Ng, Doron Shmueli, and Emily J. Shaw. 2019. *Validity of the SAT for Predicting First-Year Grades and Retention to the Second Year*. College Board.

Zimmerman, Seth D. 2014. "The Returns to College Admission for Academically Marginal Students." *Journal of Labor Economics* 32 (4): 711–54.

Zitzewitz, Eric. 2025. "The Academic Consequences of Affirmative Action Bans Combined with Diversity Targets." NBER Working Paper 33563.

Journal of Economic Perspectives—Volume 40, Number 4—Fall 2026—Pages 31–52

Can Government Support for Innovation Create the Next Silicon Valley?

E. Jason Baron, Shawn Kantor, and Alexander Whalley

Silicon Valley is the world's most prominent knowledge hub, but no one has identified the policy that created it. That absence is awkward for a growing set of programs that aim to build knowledge hubs on purpose. In this essay, we ask what public research funding can and cannot do for a regional economy. Our answer is that public research and development is one input into a knowledge hub and not necessarily the binding one. Its local returns may depend on complementary inputs. Where entrepreneurs, risk capital, and deep markets for skilled labor are already in place, public research funding can stimulate local economic activity. Where those inputs are thin, it may fund research without generating much follow-on economic activity.

The question is a live one. If research and development has helped fuel the success of today's knowledge hubs, it is natural to ask whether targeted place-based investments can replicate that success elsewhere (Gruber and Johnson 2019). One recent initiative along these lines is the National Science Foundation's Regional Innovation Engines program, authorized by the Creating Helpful Incentives to Produce Semiconductors (CHIPS) and Science Act of 2022. The program envisions "supporting multiple flourishing regional innovation ecosystems across the

■ *Jason Baron is Assistant Professor of Economics, Duke University, Durham, North Carolina. Shawn Kantor is the L. Charles Hilton Jr. Distinguished Professor of Economics, Florida State University, Tallahassee, Florida. Alexander Whalley is Professor of Economics, University of Calgary, Calgary, Alberta, Canada. Baron is a Faculty Research Fellow and Kantor is a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are jason.baron@duke.edu, skantor@fsu.edu, and alexander.whalley@ucalgary.ca. Author names are alphabetically ordered.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251493.

US, spurring economic growth in regions that have not fully participated in the technology boom of the past few decades."[1] Given growing policy interest in spatial economic inequality across the United States and in place-based approaches to addressing it, understanding the returns to this kind of spending is important.

Although defining a knowledge hub is challenging (Kerr and Robert-Nicoud 2020), we characterize it here as a geographic area with a dense concentration of knowledge, creation, and innovation. These regions typically feature universities and research laboratories, technology firms, venture capital, and a highly skilled workforce that, together, support both the production and commercialization of new ideas (Nelson 1993; Feldman 1994; Edquist 1997; Bresnahan and Gambardella 2004; Furman, Porter, and Stern 2002). The definition itself carries the central point of this essay: A hub has many inputs, and public research funding is only one of them. Five economic considerations follow from that observation, and they organize what comes next.

First, public research produces an input, not a product. To the extent that government research is basic, it complements applied private-sector research and development rather than substituting for it (Jaffe 1989; Azoulay, Graff Zivin, and Sampat 2011). Someone else has to turn the resulting knowledge into commercial activity. This complementarity is what makes local effects possible; it is also what makes those effects conditional on the presence of firms capable of putting that knowledge to use.

Second, knowledge does not travel far without cost. Frictions in diffusion generate home bias, so knowledge is easier to access near where it is produced (Ellison and Glaeser 1997; Moretti 2004b; Atkin, Chen, and Popov 2022; Liu and Su 2026). This distance decay creates an advantage for firms and workers located near the source of new knowledge. However, a local advantage need not represent the creation of new economic activity; some of the gains may instead be drawn from elsewhere. The distinction is easily overlooked in evaluations of place-based programs.

Third, the complementary inputs are frequently the scarce ones. Risk capital, experienced entrepreneurs, and a deep market for skilled labor are often critical to translating research into local commercial activity. If the technology of hub formation is multiplicative in these inputs, the marginal return to research funding is low where these complements are thin. This is the core of our argument, and it explains why identical grants can produce very different local outcomes.

Fourth, initial conditions matter because many of the complements to research are locally persistent and slow to build. There is also an element of good fortune in where hubs ultimately sprouted. Family ties influenced the location decisions of early technology pioneers in both Silicon Valley and Seattle. Boston and San Diego benefited from having scientists already in residence just as technologies relying on their expertise emerged. Under increasing returns, such accidents can push a

[1] See Guzman, Murray, Stern, and Williams (2024) for an assessment of how the Engines program can stimulate a localized innovative ecosystem, and https://new.nsf.gov/funding/initiatives/regional-innovation-engines/about-nsf-engines for program details.

region above a threshold that sets in motion a self-reinforcing process of agglomeration. Such initial conditions are difficult for policy to manufacture after the fact.

Fifth, creation and reallocation are different outcomes. A positive local estimate is consistent with a national effect of zero if the activity was drawn from somewhere else. This distinction matters for both the efficiency case for public research and development and the distributional case for place-based research and development policy, and the two can point in opposite directions.

The evidence we review bears out these insights. Public research and development can stimulate local economic activity, but its effects vary considerably across settings, are often modest, and may partly reflect the reallocation of activity from other places. Our assessment complements recent work on the role of public research and development in stimulating technological innovation (Bloom, Van Reenen, and Williams 2019) and on the broader mix of factors necessary for the formation of technology clusters (Kerr and Robert-Nicoud 2020). It also connects to work emphasizing how knowledge spillovers and agglomeration can contribute to spatial divergence (Moretti 2012).

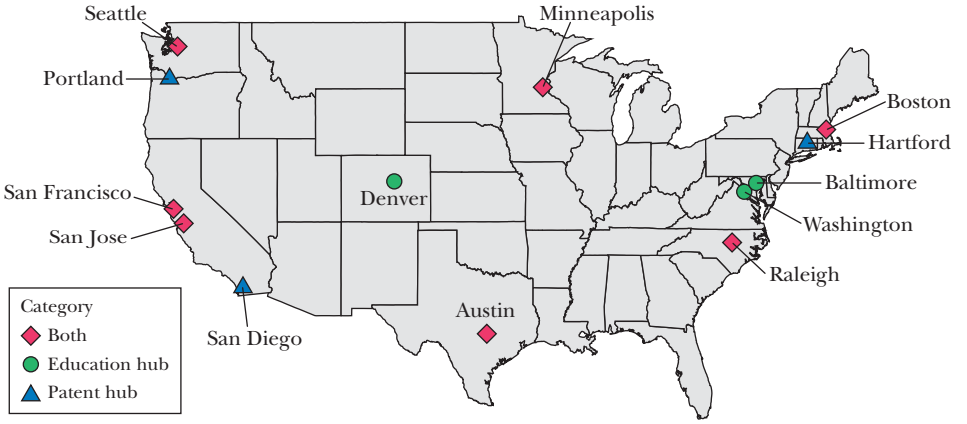
We close by discussing the experience of the University of California, Merced—the first research university established in the United States in the twenty-first century—founded in part to revitalize the local economy. Its first two decades illustrate our central point. Despite a large, deliberate, and sustained public investment in research, the university was established in a region where many of the complementary inputs to a knowledge hub were scarce, and its effects on the local economy have, so far, been modest. Building a thriving technology hub in Merced may require policies that attract or develop these complementary inputs as well.

Knowledge Hubs and Spatial Divergence

Identifying Knowledge Hubs

One method of seeking out knowledge hubs, both widely used and necessarily limited, focuses on the geographic concentration of patents per capita. Another method focuses on areas with high shares of educated workers. Here, we focus on US metropolitan statistical areas (MSAs) with population of at least 1 million, which includes roughly the 50 most populous US urban areas. The map in Figure 1 shows the urban areas that are in the top ten in either patents per capita or the percentage of the adult population with at least a bachelor's degree. Urban areas that are high in patents per capita but not education are represented by triangles, those high in education but not patents are represented by circles, and those high in both are shown with diamonds. Prominent knowledge hubs that rank highly in both patents and education include Boston, Raleigh-Durham, Austin, Minneapolis, San Francisco, San Jose, and Seattle. Additionally, Hartford, San Diego, and Portland are at the top of the distribution for patents, while Baltimore, Washington, DC, and Denver are top educational hubs.

Figure 1
Mapping Knowledge Hubs, 2015–2019



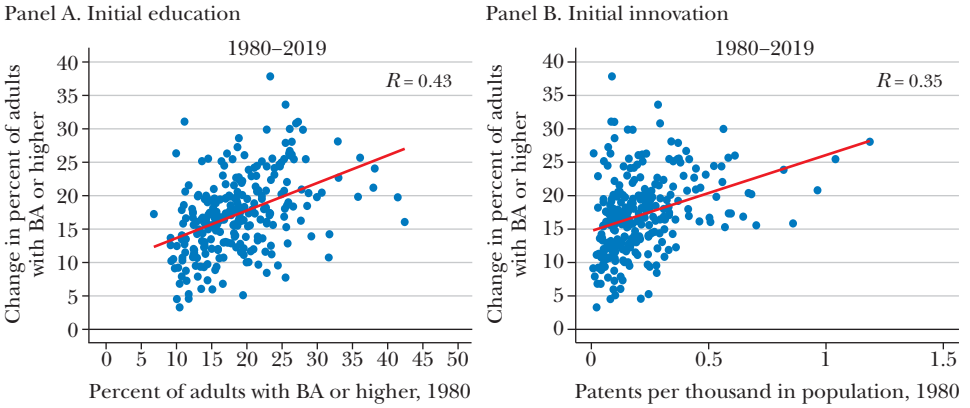
Source: Authors' calculations using American Community Survey (ACS), Integrated Public Use Microdata Series (IPUMS), National Historical Geographic Information System (NHGIS) data. The educational attainment data are county- level data from the 2015–2019 five-year American Community Survey. They were retrieved from the IPUMS database's NHGIS. See Manson et al. (2023). Data for county-level patents per thousand in population are simple means from 2015–2019 to mirror the American Community Survey years selected. The patent data are from Marco, Carley, Jackson, and Myers (2015), Petralia, Balland, and Rigby (2016), and the US Patent and Trademark Office (2024).
Notes: The samples are restricted to metropolitan statistical areas with a population of at least 1 million. The MSAs represented are those in the top ten for (1) percent of the population with a bachelor's degree or higher and (2) patents per thousand in population. If in group a only, it is labeled an education hub. If in group b only, it is labeled a patent hub. If in both groups a and b, it is labeled both. Data are aggregated from counties to MSAs. MSA counts are weighted means by population of the county counts of percent of the adult population with a bachelor's degree or higher and the number of patents per thousand in population.

What Drives Spatial Divergence?

To understand whether replicating knowledge hubs in lagging communities could drive economic growth, one must first grasp why economic success varies across regions. While the exact reasons remain debated, it is widely recognized that the concentration of skilled workers in specific areas significantly contributes to regional income differences (Benjamin, Glaeser, and Summers 2018; Glaeser and Saiz 2003; Moretti 2012; Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer 2013; Valero and Van Reenen 2019).

Counties that started with more human capital have tended to accumulate more of it over time, contributing to persistent differences in educational attainment across places. Panels A and B of Figure 2 show that counties with a higher initial share of college-educated adults and greater patenting per capita in 1980 experienced larger increases in the share of adults with a bachelor's degree or higher between 1980 and 2019. These patterns are consistent with agglomeration forces in innovative sectors, through which firms and skilled workers are drawn to locations with favorable initial

Figure 2
Initial Levels and Subsequent Growth of Educational Attainment



Source: The data are US Census county-level data from the 1980 decennial census and 2015–2019 five-year American Community Survey. They were retrieved from the Integrated Public Use Microdata Series database's National Historical Geographic Information System. See Manson et al. (2023). The patent data are from Marco, Carley, Jackson, and Myers (2015), Petralia, Balland, and Rigby (2016), and the US Patent and Trademark Office (2024).
Note: The figures draw from the sample of counties with a 2019 population of at least 250,000.

conditions, including existing concentrations of human capital and established networks (Moretti 2012). Recent evidence also suggests that very local productivity spillovers coexist with substantial sorting of higher-quality firms into more productive locations (Baum-Snow, Gendron-Carrier, and Pavan 2024).[2]

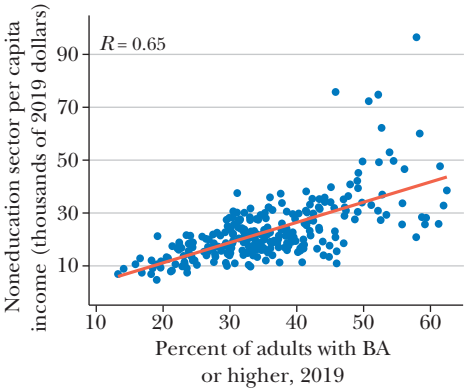
The focus on human capital reflects the understanding that in a knowledge-based economy, the level of human capital (or education) within a region is a strong predictor of economic prosperity. Figure 3 shows a clear positive relationship between the share of educated workers in a county (panel A) or its innovative capacity (as proxied by patent production per capita in panel B) and its income in noneducation sectors in 2019. Ultimately, the stock of human capital in a region or area plays a vital role in its success in an innovation-driven economy.

Although the correlations in Figure 3 do not establish causality, they are consistent with a broader literature linking human capital to regional economic prosperity. Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer(2013), using data from 1,569 regions across 110 countries, find that human capital is particularly important in accounting for regional differences in development. More direct evidence of a causal effect comes from Moretti (2004a), who instruments for a city's share of college graduates using its demographic structure and the presence of a land-grant college and

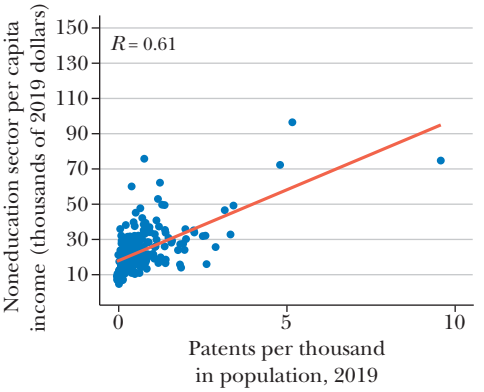
[2]Almazan, De Motta, and Titman (2007) show, theoretically, that firms optimally choose to locate in a "cluster" when workers bear training costs and firm-specific productivity shocks can be large. In this setting, firms benefit from access to a pool of workers with embodied, industry-specific human capital. Conversely, when firms bear substantial training costs, they have an incentive to locate in more isolated locations.

Figure 3
Relationship between Human Capital and Innovation and Noneducation Sector Income

Panel A. Relationship between the percent of the adult population with a BA and noneducation sector income



Panel B. Relationship between the number of patents per thousand population and noneducation sector income



Sources: Authors' calculation using data from the American Community Survey, Integrated Public Use Microdata Series National Historical Geographic Information System, Marco, Carley, Jackson, and Myers et al. (2015), Petralia, Balland, and Rigby (2016), and US Patent and Trademark Office (2024). The non-education sector income is from US Census Bureau (2019). The educational attainment data are county- level data from the 2015–2019 five-year American Community Survey. They were retrieved from Integrated Public Use Microdata Series database's National Historical Geographic Information System. See Manson et al. (2023). The patent data are from Marco, Carley, Jackson, and Myers et al. (2015), Petralia, Balland, and Rigby (2016), and US Patent and Trademark Office (2024).
Note: The figures draw from the sample of counties with a 2019 population of at least 250,000.

finds that increases in the local share of college graduates raise wages across education groups. Glaeser (2005) provides complementary historical evidence on the role played by local human capital in Boston's sustained economic growth.

Creating knowledge hubs in lagging communities cannot occur in a vacuum, as a successful ecosystem depends on multiple interconnected factors. Sorting out the necessary from the sufficient conditions for success is therefore a critical first step. There seems little doubt that human capital is an important building block of a new knowledge hub.

Motivation for Knowledge-Based, Place-Based Policies

As Bartik (2020) notes, "Places matter for policy because places matter to people." Given the disparity in economic performance across regions, one might question why more people do not simply move to opportunity and prosperity, as economic models might predict. After all, historical American economic success has relied on such population mobility among its own citizens and, indeed, among those from around the world willing to migrate. However, migration is not frictionless. Accessing US regions that already have knowledge hubs poses a significant challenge due to high costs of living, especially driven by high housing prices. These

areas, known for their robust job markets, often contend with strict local zoning laws and regulations that limit land use, impeding the construction of sufficient housing to meet increasing demand. The considerable cost of living in most knowledge hubs remains a formidable barrier.

Furthermore, highly educated Americans are more likely to move than those with lower levels of education, exacerbating economic inequalities among those who are educated and those who are less so. This trend further consolidates income in a few affluent areas, leaving behind those who cannot afford or opt not to move. These challenges underscore the rationale for place-based policies aimed at replicating existing knowledge hubs and revitalizing economically disadvantaged communities.

The Role of Luck in the Rise of Knowledge Hubs

Many well-established knowledge hubs today owe at least part of their success to a considerable dose of luck, whether through the serendipitous location decisions of key entrepreneurs and inventors or the presence of academics working on emerging technologies at opportune moments (Moretti 2012). It is worth examining some of these success stories. An important caveat, however, is that these cases are selected precisely because they became successful knowledge hubs and therefore need not be representative of the typical experience.

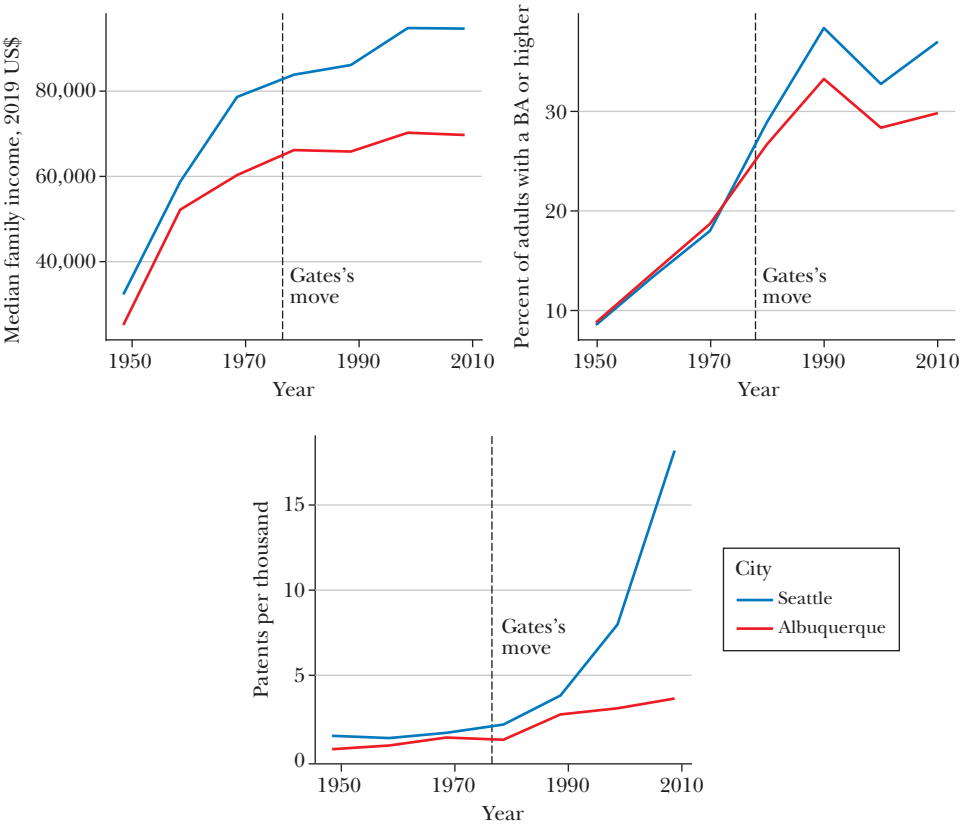
Family-Driven Serendipity

Consider the role that Microsoft has played in Seattle's economic success. Microsoft was founded in 1975 in Albuquerque, New Mexico. Despite initial success in Albuquerque, with year-end sales exceeding $1 million by 1978, Microsoft's founders Bill Gates and Paul Allen were drawn back to their roots in Seattle, where they grew up and attended the same high school. Gates and Allen decided to move the company to Bellevue, across Lake Washington from Seattle (NPR 2008). In 1979, Seattle was not an obvious choice for a software company. It resembled Detroit more than Silicon Valley, with an economy reliant on declining manufacturing, together with low levels of education and high crime rates (Moretti 2012).

Microsoft's move provides an illustrative example of how a seemingly idiosyncratic location decision can precede substantial changes in a local economy. Figure 4 compares the subsequent evolution of Seattle, which gained Microsoft in 1979, with Albuquerque, which lost it. Before the move, the two cities were relatively similar in their shares of college-educated adults and patents per capita, although median family income was already substantially higher in Seattle. After the move, the gaps widened substantially. Seattle experienced larger increases in educational attainment and median family income, while patenting per capita surged relative to Albuquerque.

Silicon Valley has a similar foundational story involving the serendipitous relocation of William Shockley, who shared the 1956 Nobel Prize in Physics for work leading to the invention of the transistor. Dissatisfied with his career progression at Bell Labs in Murray Hill, New Jersey, Shockley returned to Palo Alto in 1956,

Figure 4
Microsoft Serendipity: Seattle Versus Albuquerque

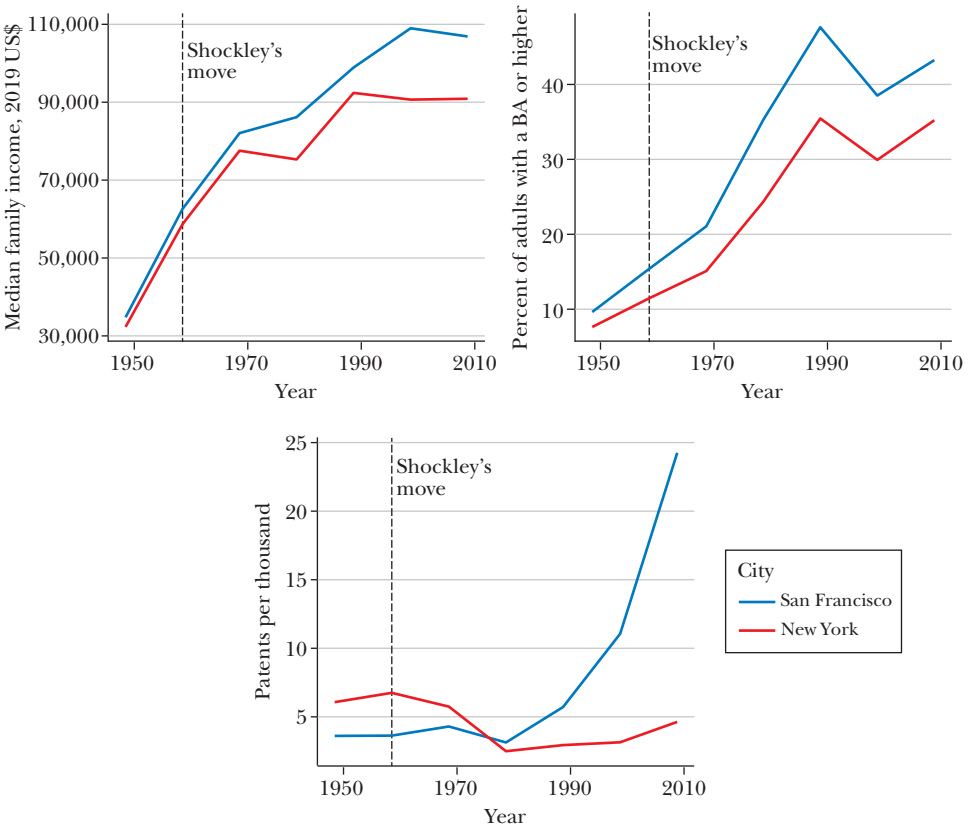


Source: Authors' calculation using data from United States Census, Integrated Public Use Microdata Series National Historical Geographic Information System,Marco, Carley, Jackson, and Myers et al. (2015), Petralia, Balland, and Rigby (2016), and US Patent and Trademark Office (2024). The educational attainment and family income data are US census county-level data cross-walked to metropolitan statistical area. They were retrieved from IPUMS database's NHGIS. See Manson et al. (2023). The patent data are from Marco, Carley, Jackson, and Myers et al. (2015), Petralia, Balland, and Rigby (2016), and US Patent and Trademark Office (2024).

where his ailing mother resided (Markoff 2022). Fred Terman, Stanford University's provost and a key figure in the rise of Silicon Valley, offered Shockley space in Stanford's new industrial park, one of the first university-led research parks and an important building block in the region's emerging technology cluster. Shockley founded the Shockley Semiconductor Laboratory in 1956. Despite facing initial recruitment challenges, his lab became a critical catalyst in the emergence of Silicon Valley (PBS 1999). Thus, while Stanford University played a role after Shockley moved, it was not the main determinant of Shockley's move.[3]

[3]As Moretti (2012, p. 186) points out, "A research university was necessary but far from sufficient for the birth and coming-of-age of the Valley. If Shockley had decided to locate in, say, Providence, which

Figure 5
Shockley Serendipity: San Francisco Versus New York



Source: Authors' calculation using data from United States Census, Integrated Public Use Microdata Series National Historical Geographic Information System, Marco, Carley, Jackson, and Myers et al. (2015), Petralia, Balland, and Rigby (2016), and US Patent and Trademark Office (2024). The educational attainment and family income data are US census county-level data cross-walked to metropolitan statistical area. They were retrieved from IPUMS database's NHGIS. See Manson et al. (2023). The patent data are from Marco, Carley, Jackson, and Myers et al. (2015), Petralia, Balland, and Rigby (2016), and US Patent and Trademark Office (2024).
Note: This figure compares the New York City metropolitan statistical area to the San Francisco metropolitan area. Shockley worked at the Bell Labs facility located in Murry Hills, New Jersey, which is located in the New York City MSA.

Figure 5 displays the time series of median family income in the San Francisco and New York metropolitan statistical areas (the latter which includes Murray Hill, New Jersey), before and after Shockley's relocation to the San Francisco Bay Area. Initially, both areas had nearly identical median family incomes. However,

was then an area with a significantly more developed industrial base than Palo Alto, Silicon Valley might today be clustered in Rhode Island, and we would be reading dozens of books on how Brown University caused the cluster."

following Shockley's move in 1956, their economic paths diverged significantly, with San Francisco experiencing a notably greater increase in median income compared to New York, especially after about 1990. The right-hand panel of the figure shows a similar divergence in human capital, as measured by the percentage of adults with a bachelor's degree or higher. Moreover, the share of patents over time surged in San Francisco in the decades following Shockley's arrival, despite roughly similar trends in both areas prior to his move. Silicon Valley therefore appears to owe at least part of its emergence to Shockley's decision to locate near his family.

Although family connections surely do not fully explain the booms in Seattle or Silicon Valley, neither example gives public research and development a central role in determining where the hub initially formed. In Seattle, Microsoft's location decision was largely unrelated to universities or public research, while in Silicon Valley, Stanford provided important complementary infrastructure after Shockley had chosen to return to the region. These histories reinforce the distinction between creating the initial conditions for a hub and providing complementary inputs that can amplify them once they are in place.

Technology-Based Serendipity

Another type of serendipity that has contributed to the establishment of high-profile hubs in the United States is the presence of key academics focusing on emerging technologies. This is evident in the emergence of biotechnology hubs in the 1970s.

Boston-Cambridge, the San Francisco Bay Area, and San Diego stand out as the top three areas with the densest clusters of biotech firms today. As Moretti (2012) points out, while it is often believed that the strong presence of renowned universities like MIT, Harvard, Stanford, University of California, Berkeley, and University of California, San Diego in these regions played a pivotal role in nurturing biotech clusters, simply having prestigious universities is clearly not sufficient to drive significant biotech activity. For example, cities such as New Haven, Connecticut (home to Yale University) and Ithaca, New York (home to Cornell University) also boast prestigious universities and top-notch biotech departments, but lack a high concentration of local biotech firms today.

The rise of major biotech hubs in the United States can be attributed in part to serendipitous factors related to specific academics pursuing research linked to emerging technologies (Zucker, Darby, and Brewer 1998). Following the pivotal discovery of recombinant DNA in the early 1970s—a breakthrough important for the production of pharmaceuticals, enzymes, and other biotechnological products—the incentives for biotech research skyrocketed. Over the subsequent 15 years, the previously nascent US biotech industry grew to 700 active firms.

What led these firms to congregate in specific areas? While intellectual human capital thrived around prestigious universities, the presence of exceptional scientists, as measured by research productivity, played a distinct and significant role beyond the mere existence of these universities and their associated research

funding (Zucker, Darby, and Brewer 1998, p. 302). In essence, the success of biotech hubs can be attributed to the initial presence of academic "stars" in biotech, such as Herbert Boyer, Stanley Cohen, and Paul Berg in the Bay Area. Boyer, who was a professor at the University of California, San Francisco, co-founded Genentech. In the Boston area, Walter Gilbert, a cofounder of Biogen, was a professor at Harvard and won the 1980 Nobel Prize in Chemistry for his work on DNA sequencing. Phillip Sharp, a molecular biologist at MIT, who won the Nobel Prize in Physiology or Medicine in 1993 for discovering RNA splicing, was a cofounder of Biogen and later Alnylam Pharmaceuticals. These academic stars helped attract a wave of new biotech enterprises eager to leverage their scientific research.

Figure 6 illustrates the time series of median family income in Boston, San Diego, and the San Francisco Bay Area—today's top three biotech clusters—compared with eight other metropolitan areas associated with universities that had top-15 biochemistry departments in 1966: Baltimore, Maryland; Champaign, Illinois; Cleveland, Ohio; Durham-Chapel Hill, North Carolina; Los Angeles, California; Madison, Wisconsin; New York City, New York; and Seattle, Washington. These two groups of urban areas had similar median family incomes up until 1980, after which their economic trajectories diverged. Today, the top three biotech clusters boast a median family income roughly $15,000 higher than the others. This pattern is also evident in the percentage of adults with a bachelor's degree and the number of patents per capita. Overall, the evidence presented in the figure is consistent with the hypothesis in Zucker, Darby, and Brewer (1998) that prominent knowledge hubs have emerged from the serendipitous colocation with star scientists with expertise in critical emerging new technologies.

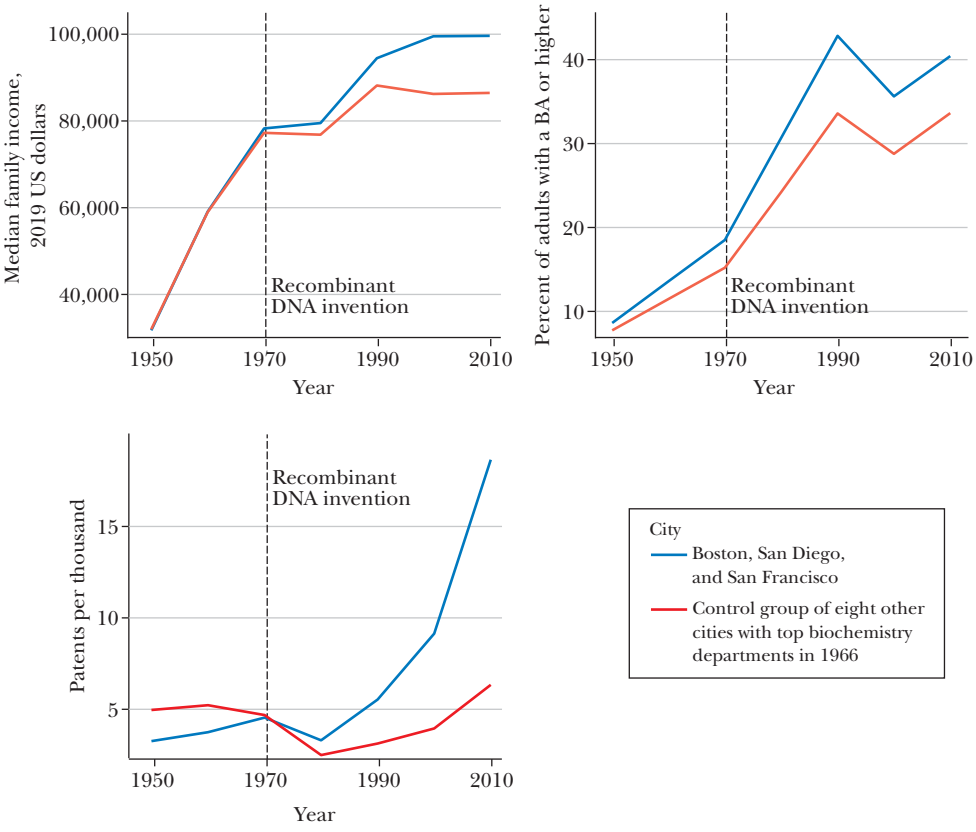
The pattern of serendipity described in this section extends to industries beyond knowledge-based technology hubs. For example, similar arguments can be made regarding the rise of Hollywood, linked to the pioneering director D.W. Griffith (Moretti 2012). When an industry has been successful in a given location for a long period, there can be a tendency to presume that the location was somehow predestined to become a hub for that industry. However, their origins often trace back to serendipitous events or the influence of key visionary individuals who ignited their growth.

While we have a clear understanding of what happens once a hub is established and agglomeration forces generate their virtuous cycle of economic prosperity, predicting the emergence of many tech hubs remains challenging due to the random nature of their beginnings and the fact that in retrospect they seem quite similar to "runner-up" places.

The Role of Policy Intervention in the Rise of Technology Hubs

While many well-known knowledge hubs owe a significant share of their success to sheer luck, is there empirical evidence that policymakers can directly influence their creation as well? This section examines the empirical evidence for a commonly

Figure 6
DNA Luck: Growth of Biotech Versus Non-Biotech Hubs



Sources: Authors' calculation using data from United States Census, Integrated Public Use Microdata Series National Historical Geographic Information System, Marco, Carley, Jackson, and Myers et al. (2015), Petralia, Balland, Rigby (2016), and US Patent and Trademark Office (2024) and Carrter (1966) university rankings data. The educational attainment and family income data are US census county-level data cross-walked to metropolitan statistical area. They were retrieved from IPUMS database's NHGIS. See Manson et al. (2023). The patent data are from Marco, Carley, Jackson, and Myers et al. (2015), Petralia, Balland, and Rigby (2016), and US Patent and Trademark Office (2024).
Note: This figure compares three leading contemporary biotech hubs—Boston, San Diego, and San Francisco—with other metropolitan statistical areas containing one or more of the top 15 biochemistry departments in 1966. Boston and San Francisco themselves contained top 15 departments; San Diego is included in the treatment because it subsequently emerged as a leading biotech cluster. The control group includes: Baltimore, Maryland; Champaign, Illinois; Cleveland, Ohio; Durham-Chapel Hill, North Carolina; Los Angeles, California; Madison, Wisconsin; New York City, New York; and Seattle, Washington.

proposed policy to nurture a knowledge hub: increased public research and development investment.

Public research and development funding has played important roles in driving knowledge-based economic growth since the second half of the twentieth century,

spurring technological advancements, bolstering key industries, improving public health, and establishing some of the building blocks for today's knowledge-based economy. US federal research and development funding as a percentage of GDP has fluctuated over the decades, but significant investment episodes occurred in the mid-twentieth century, especially during World War II, the Space Race, and the Cold War. Today, however, federal research and development funding as a share of gross domestic product is a fraction of its historical peak.[4]

Of course, the case for public support of research and development is that technological innovation—perhaps especially when aimed at basic research and development—has broader spillovers. Indeed, the larger the geographic region, the more that the positive externalities from such research and development can be internalized. Here, however, we are less interested in the broad case for public support of research and development and instead focus on whether it can "jump-start" economically lagging communities. As proposed in Gruber and Johnson (2019), for example, the idea is for public research and development funding to complement local private-sector investment. In the discussion that follows, we consider the evidence that an expansion of public research and development in a particular area can serve as a place-based means to stimulate local economic growth. We organize the discussion around the source of exogenous shocks to research and development funding, which allow this research to draw plausibly causal inferences.

Short-Term Effects: Endowment Shocks at Universities

We begin by reviewing evidence on the short-term effects of research conducted at universities, an important component of public research and development in many parts of the country. Kantor and Whalley (2014) study noneducation-sector wage growth within urban counties, examining how short-term changes in the size of the research university sector relate to wage and labor productivity growth locally.

As research university activities are certainly not random and may depend on support from local sources, university expenditure itself may reflect unobserved performance of the local economy. To obtain a causal estimate of how research university efforts affect labor productivity in other sectors, the authors utilize the largely fixed share of university spending from endowment holdings to generate exogenous shocks to the availability of annual funding. To the extent that universities typically spend between 4 to 5 percent of their respective endowments each

[4]For example, focusing on the National Aeronautics and Space Administration (NASA), specifically, its budget reached a peak of 4.4 percent of all federal spending in 1966. The agency's budget did decline after the landing on the moon was successfully accomplished in 1969, but still accounted for 1.92 percent of federal spending in 1970. Subsequently, the level of spending fluctuated between 0.75 and 1 percent of the federal budget from 1975 until the end of the twentieth century. To provide some perspective on the magnitude of NASA's budget during the Space Race and some understanding of the decline in federal research and development, consider that in 2020 the total of all nondefense federal research and development amounted to 1.5 percent of the federal budget.

year, fluctuations in capital market returns will manifest in different exogenous amounts of spending available to those universities with large endowment holdings versus smaller holdings.

These exogenous increases in university research activity result in productivity spillovers to other industries: Specifically, a 1 percent increase in university expenditures in a county increases local labor income in other sectors by 0.08 percent. Bergeaud et al. (2025) show that public research funding in France crowds in local private research and development for roughly one-to-one (€0.81 per €1) in scientifically proximate industries, mainly through formal public–private contracting. This effect persists for at least five years, which suggests that the impact of research expenditures is something more than a short-run stimulus to local labor demand. Putting this estimate into some relative perspective, however, they find that the elasticity of the effect is no greater than those accruing from the agglomeration benefits of population growth. The effects of short-term fluctuations in research spending may be smaller than the effects of long-term increases in research spending when follow-on investments occur with a lag. The effects of short-term government funding shocks may also differ from short-term endowment-based shocks. Babina et al. (2023) show that when federal research funding falls, industry funding fills the gap but redirects outputs toward patents and careers appropriated by incumbent firms and away from researcher-led startups.

Short-Term Effects: A Shift in Legal Regime

Before 1980, universities lacked strong incentives to commercialize research, because the federal government held default rights to intellectual property produced by universities under federally funded research. However, the 1980 Bayh-Dole Act conveyed property rights over innovations that universities developed under the auspices of federal funding. Hausman (2022) examines the effect of this shock, which radically changed the potential local spillovers associated with university research activity, by exploiting cross-sectional variation in the extent to which industries were technologically related to nearby university innovation—as measured by patents—and were, thus, more likely to benefit from increased access to its ideas after Bayh-Dole. The exogenous nature of the policy change facilitates the identification of local effects of university research commercialization.

Using firm-level micro data, Hausman shows that employment, payroll, and wages grew differentially faster after the Bayh-Dole Act in industries more closely related to the technological strengths of nearby universities. When contextualizing the magnitude of these findings, however, Hausman notes that the magnitudes of the estimated effects are remarkably similar to the impacts of empowerment zones found in other work (for example, Corinth and Feldman 2024).

Consistent with our earlier discussion, changes to human capital levels were an important driver of this effect. Specifically, Hausman (2022) finds that "the growth effect is driven by nearby entry in university-linked industries, especially of multi-unit expansions; these firms disproportionately partner with universities in R&D, transfer IP [intellectual property], and innovate."

Long-Term Effects: Establishing New Institutions

What about long-term effects? Is it possible that the economic impact of local innovation activity unfolds gradually, eventually attracting the investment required to create a regional advantage that fosters the development of an enduring knowledge hub? Evidence from the establishment of the Diamond Light Source, a particle accelerator facililty in the United Kingdom, shows that proximity to a major scientific facility increased related research output, primarily by attracting additional research inputs rather than by increasing the productivity of existing scientists (Helmers and Overman 2017).

For a modern reader, American agriculture may seem an unlikely setting in which to study technology hubs. Yet agriculture offers two important advantages for measuring the long-run local effects of knowledge production. First, county-level data on farm productivity extend back roughly 150 years, allowing researchers to trace local productivity changes over an unusually long period. Second, the Hatch Act of 1887 led to the establishment of agricultural experiment stations at land-grant universities, sharply expanding local research activity in treated places. This institutional change created plausibly exogenous variation in local exposure to agricultural innovation, allowing researchers to compare productivity before and after the establishment of experiment stations and across places with and without nearby stations. Agricultural experiment stations are not knowledge hubs in the modern sense used elsewhere in this essay, but they provide unusually long-run evidence on a mechanism central to knowledge hubs: whether proximity to publicly funded research raises local productivity and how those benefits evolve over time.

Kantor and Whalley (2019) show that beneficial localized research proximity effects started out small and grew, peaking between 20 and 30 years after the experiment stations opened. While research expenditures continued to grow, research proximity effects declined, however, dying out by 50 years after the stations' opening. Thus, the expansion of university agricultural research led to a long-term but not permanent increase in local productivity.

Part of the explanation for the gradual diminution of productivity gains is that structural changes in the US economy and technological changes, such as the widespread use of automobiles and telephones, meant that ideas were able to spread more widely and faster by World War II. As new innovative ideas could spread more easily to more distant farmers, the relative benefits associated with having research conducted locally diminished. Similar changes over time in the structure of agricultural innovation toward large private sector players also likely contributed. The diminishing size of the local effects across economic eras suggests that complementary factors that have changed over time may be important in understanding knowledge hubs.

When the land grant universities were established, there was typically a documented narrative of possible site locations. Using the historical record, Andrews (2023) can compare the university locations against a set of counterfactual runner-up locations that were strongly considered for a university but ultimately lost out. His results show that establishing a college caused 62 percent more patents

per year for the area receiving the university. The effects are not primarily driven by university alumni or faculty (the "direct" effects), however. Instead, "indirect" effects dominated. Establishing a college caused a roughly 65 percent increase in county population relative to the runner-up counties, suggesting that much of the increase in total patenting may reflect the growth of the local population rather than an increase in patenting per capita.

Moonshots

Large, geographically concentrated public research and development shocks leave a durable mark on local innovation. Their effect on local economic prosperity can be more modest, and a substantial share of the local gain is drawn from other regions rather than created. Two large historical episodes provide the cleanest evidence, because in both cases the money arrived for reasons that had little to do with the economic prospects of the places that received it.

Wartime Mobilization.—Gross and Sampat (2023) show that a windfall of federal applied research funding during World War II persistently raised local innovation. Studying the Office of Scientific Research and Development (OSRD), one of the largest public investments in applied research and development in US history, they find that patenting rose in firms that received OSRD contracts and also in local firms that did not, that these effects persisted and grew over the following decades, and that they extended downstream to employment and firm creation. OSRD funding was central to the postwar emergence of electronics and microwave engineering along Route 128 near Boston and of communications and electronics in central New Jersey and greater New York.

The research design harnesses the military role in allocating research funding. The Office of Scientific Research and Development allocated contracts under acute wartime urgency, to whichever laboratories and firms already possessed the relevant technical capability, in whichever technologies the military needed at that moment. Allocation was therefore driven by short-run military demand and by the accident of where relevant expertise happened to sit in 1940, not by any judgment about the postwar commercial promise of a place. The authors exploit this by comparing technology-by-county cells that received funding to cells that did not, holding constant prewar innovative capacity in the same technology and place. The identifying assumption is not that funded and unfunded places were alike but that conditional on prewar capability, which technologies were pulled into the war effort was orthogonal to their later civilian trajectory. Two features make this credible. First, the shock was abrupt and its timing set abroad. Second, within places that had comparable prewar capability, funding tracked the technical requirements of the war rather than local economic conditions.

The Space Race.—Kantor and Whalley (2025) show that a second, large windfall raised local manufacturing activity, but not local productivity, and that its aggregate return was below that of government spending more broadly. The Soviet launch

of Sputnik in 1957 precipitated a geopolitical crisis, the creation of the National Aeronautics and Space Administration in 1958, and a national commitment to a crewed lunar landing. At its peak in the mid-1960s, NASA absorbed over 0.7 percent of gross domestic product and the program employed more than 400,000 workers. They find that space race spending had a persistent positive effect on manufacturing value added, employment, and capital investment in space relevant industries in exposed locations, but there were no detectable productivity gains, either in the space-related sectors themselves or among firms colocated near them.

The variation comes from technological capability that predated the shock. Certain locations and industries had developed building blocks before Sputnik that the lunar program would later require, and these places were mechanically more exposed to the subsequent surge in federal spending. Interacting this predetermined local composition with the national escalation in the National Aeronautics and Space Administration outlays yields exposure that varies across places for reasons fixed before anyone knew the shock was coming. The credibility of the design rests on the source of the shock rather than on the shares: The escalation was driven by a foreign launch and the geopolitical response to it, not by the economic conditions of any American region. This matters because the siting of major NASA centers was a political decision, and using it directly would confound the effect of research spending with the characteristics that attract congressional influence. Keying on pre-Sputnik technological capability sidesteps that problem.

Three further results bear directly on the question of this essay. First, the estimated effects, while economically meaningful, are modest relative to their cost. The implied fiscal multiplier was only 0.3 to 0.4. By comparison, Ramey's (2019) survey of the macroeconomic literature concludes that most estimates of government spending multipliers, for broad spending categories averaged over samples, fall between 0.6 and 0.8 or perhaps as high as 1, and local fiscal multipliers are larger still (Chodorow-Reich 2019). Second, no clear productivity gains emerged, in the space industries or among their neighbors, which is precisely the channel through which a knowledge hub provides particularly large benefits to the local real economy. Third, inventor migration data show that the National Aeronautics and Space Administration spending pulled inventors toward newly funded locations. When the Manned Spacecraft Center opened in Houston in 1963, nearly all technical and clerical staff could be hired locally, but only about 10 percent of the 6,000 scientists, engineers, and administrators came from the Houston area (Holman and Konkel 1968). Because NASA contracted heavily to the private sector, more than 125 firms with a presence in the space field opened Houston offices within five years, including General Electric, Honeywell, IBM, North American Aviation, Lockheed, Raytheon, Texas Instruments, and TRW (Brady 2007). This beggar-thy-neighbor effect completely offset whatever benefits neighboring regions might have realized through knowledge spillovers or increased demand that flowed nationally.[5] The location of space innovation hubs has

[5] Estimating spatial spillovers allows nationwide and local impacts to be compared through a spatial equilibrium model. While Kantor and Whalley (2025) do not find large nationwide effects, recent work using

proven remarkably persistent, with hubs from the Cold War remaining dominant in the SpaceX era (Gaetani and Whalley 2026).

Conclusion

Our analysis highlights several key takeaways about the factors shaping knowledge hubs and the potential role of public research and development. First, many of today's successful and exemplary hubs owe their origins, at least in part, to a hefty dose of good luck, illustrating the unpredictable nature of their emergence.

Second, while some evidence suggests that public research and development investments can generate positive impacts in the broader economy, including sectors outside the targeted ones, the results are mixed, with spillovers varying between contexts.

Third, even when statistically significant impacts are observed, assessing their economic significance, particularly the cost-effectiveness of the public expenditure relative to other public expenditure programs, remains challenging. The few papers noted above that conduct a detailed calculation of the cost-effectiveness of research and development policy expenditures find that the magnitude of the benefits is often modest—and not larger than that of alternative interventions. In some cases, the potential local gains may be in whole or in part a reallocation of firms or workers from other areas.

Taken together, the existing evidence does not make a case that public research and development is sufficient for fostering knowledge hubs or driving large-scale economic transformation in cost-effective ways. Instead, knowledge hubs or tech clusters are ecosystems that may require synergistic interactions between knowledge producers, users, and funders before increasing returns set in (Lerner 2009).

This insight suggests that quantifying policy complementarity is crucial. For example, complementary factors that enhance research and development returns such as entrepreneurial human capital and social networks are central to Saxenian's (1994) analysis of Silicon Valley's success, yet their causal effects remain largely unknown. Investigating whether substantial complementarities exist between the strength of a preexisting local industrial ecosystem and additional research and development spending is a frontier question in this literature.[6]

---

macroeconomic narrative identification finds substantial nationwide returns to public research and development (Fieldhouse and Mertens, forthcoming), particularly for public-private partnership research and development (Gazzani, Martinez, Natoli, and Surico 2026) with spillovers manifesting in locations far from those with publicly funded research (Meyers and Lanahan 2022) and with large effects in a developing country context (Akerman, Moscona, Pellegrina, and Sastry 2025). Understanding why local and national returns diverge remains an open question.

[6] In an exploration of complementarities in a different context, Johnson and Jackson (2019) examined dynamic complementarities between two key childhood interventions, Head Start and public school funding, and found that Head Start's benefits were amplified when followed by better-funded schools, and that increases in K-12 spending were more effective when preceded by Head Start participation.

Theoretically, there are strong reasons to expect such complementarities in the context of public research and development. Areas with established innovation ecosystems—including universities, research institutions, advanced research and development infrastructure like laboratories and research parks, tech startups, and venture capitalists—are better positioned to leverage the outputs of additional public research and development investments. Moreover, in areas where there is high demand for innovative solutions from consumers or businesses, research and development investments can more rapidly lead to commercialization.

Some empirical evidence, albeit limited, supports this hypothesis. Lerner, Manley, Stein, and Williams (forthcoming) estimate that 15–25 percent of the geographic variation in commercial spillovers from university-based research is attributable to place-specific factors. Meanwhile, Kantor and Whalley (2014) find that the average spillover effect of increases in university research is modest, but knowledge spillovers were significantly larger for industries sharing a labor market with universities or for industries that utilize knowledge produced at universities more intensively. Specifically, the authors provide evidence that industries that pool labor with universities have approximately twice the responsiveness to university activity when compared to the low-pooling intensity industries. These results are also consistent with work in Ellison, Glaeser, and Kerr (2010), Greenstone, Hornbeck, and Moretti (2010), and Moretti (2004b).

A recent example illustrates the importance of complementarities for public research and development investments to generate substantial local effects. In 2005, California opened a new University of California campus on a literal greenfield site in the small city of Merced. Over the first 15 years of its existence, the state invested more than $2 billion in the University of California, Merced, in part with the goal of helping transform the economy of California's impoverished Central Valley.

Before the campus opened, Merced had few of the nonuniversity factors that characterize successful knowledge hubs. Educational attainment lagged behind other regions of California, few technology-based entrepreneurs or firms were active in the area, and venture capital was scarce. To date, the university appears to have had only a modest impact on the local economy, operating primarily through increased local employment, as documented by Lee (2019). Although job creation was significant in the service sector, there is little evidence of comparable employment gains in manufacturing or other high-skilled sectors.

It is still too soon to know the long-run impact of the University of California, Merced. According to a former campus chancellor, such investments should be viewed as well worth the expense because, "In the long run, the inventions, innovations and entrepreneurial spinoffs they tend to generate typically far outpace the initial dollars invested" (University of California, Merced 2014). It may simply take time for the complementary factors needed to transform Merced into a knowledge hub to accumulate. More likely, however, complementary policies and investments will be needed for the Merced region to realize greater benefits from the first American research university established in the twenty-first century, but the opportunity costs of bringing the University of California, Merced, to fruition are real. In a

resource-constrained political economy, the relevant question is not simply whether such investments eventually generate benefits, but whether those benefits justify the public resources diverted from other places and sectors that might have produced greater returns.

■ *The authors are grateful to the JEP editors for their many helpful comments and suggestions. We thank Spencer McCloy for his outstanding research assistance as the article was being developed.*

References

Almazan, Andres, Adolfo De Motta, and Sheridan Titman. 2007. "Firm Location and the Creation and Utilization of Human Capital." *Review of Economic Studies* 74 (4): 1305–27.

Andrews, Michael J. 2023. "How Do Institutions of Higher Education Affect Local Invention? Evidence from the Establishment of US Colleges." *American Economic Journal: Economic Policy* 15 (2): 1–41.

Atkin, David, M. Keith Chen, and Anton Popov. 2022. "The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley." NBER Working Paper 30147.

Akerman, Ariel, Jacob Moscona, Heitor S. Pellegrina, and Karthik Sastry. 2025. "Public R&D Meets Economic Development: Embrapa and Brazil's Agricultural Revolution." NBER Working Paper 34213.

Azoulay, Pierre, Joshua S. Graff Zivin, and Bhaven N. Sampat. 2011. "The Diffusion of Scientific Knowledge across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine." NBER Working Paper 16683.

Babina, Tania, Alex Xi He, Sabrina T. Howell, Elisabeth Ruth Perlman, and Joseph Staudt. 2023. "Cutting the Innovation Engine: How Federal Funding Shocks Affect University Patenting, Entrepreneurship, and Publications." *Quarterly Journal of Economics* 138 (2): 895–954.

Bartik, Timothy J. 2020. "Using Place-Based Jobs Policies to Help Distressed Communities." *Journal of Economic Perspectives* 34 (3): 99–127.

Baum-Snow, Nathaniel, Nicolas Gendron-Carrier, and Ronni Pavan. 2024. "Local Productivity Spillovers." *American Economic Review* 114 (4): 1030–69.

Benjamin, Austin, Edward Glaeser, and Lawrence Summers. 2018. "Saving the Heartland: Place-Based Policies in 21st-Century America." *Brookings Papers on Economic Activity* 49 (1): 151–232.

Bergeaud, Antonini, Arthur Guillousouic, Emeric Henry, and Clément Malgouyres. 2025. "From Public Labs to Private Firms: Magnitude and Channels of Local R&D Spillovers." *Quarterly Journal of Economics* 140 (4): 3233–82.

Bloom, Nicholas, John Van Reenen, and Heidi Williams. 2019. "A Toolkit of Policies to Promote Innovation." *Journal of Economic Perspectives* 33 (3): 163–84.

Brady, Kevin M. 2007. "NASA Launches Houston into Orbit: The Political, Economic, and Social Impact of the Space Agency on the Southeast Texas, 1961–1969." In *Societal Impact of Spaceflight*, edited by Steven J. Dick and Roger D. Launius, 451–65. National Aeronautics and Space Administration.

Bresnahan, Timothy, and Alfonso Gambardella, eds. 2004. *Building High-Tech Clusters: Silicon Valley and Beyond*. Cambridge University Press.

Cartter, Allen M. 1966. *An Assessment of Quality in Graduate Education*. American Council on Education.

Chodorow-Reich, Gabriel. 2019. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *American Economic Journal: Economic Policy* 11 (2): 1–34.

Corinth, Kevin, and Naomi Feldman. 2024. "Are Opportunity Zones an Effective Place-Based Policy?" *Journal of Economic Perspectives* 38 (3): 113–36.

Edquist, Charles, ed. 1997. *Systems of Innovation: Technologies, Institutions, and Organizations*. Routledge.

Ellison, Glenn, and Edward L Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* 105 (5): 889–927.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review* 100 (3): 1195–213.

Feldman, Maryann P. 1994. *The Geography of Innovation*. Kluwer Academic Publishers.

Fieldhouse, Andrew, and Karel Mertens. Forthcoming. "The Returns to Government R&D: Evidence from US Appropriations Shocks." *American Economic Review*.

Furman, Jeffrey L., Michael E. Porter, and Scott Stern. 2002. "The Determinants of National Innovative Capacity." *Research Policy* 31 (6): 899–933.

Gaetani, Ruben, and Alexander T. Whalley. 2026. "Old Space, New Space: A Commercial Revolution in Innovation?" NBER Working Paper 35212.

Gazzani, Andrea Giovanni, Joseba Martinez, Filippo Natoli, and Paolo Surico. 2026. "The Public Origins of American Innovation." Bank of Italy Temi di Discussione 1521.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2013. "Human Capital and Regional Development." *Quarterly Journal of Economics* 128 (1): 105–64.

Glaeser, Edward L. 2005. "Reinventing Boston: 1630–2003." *Journal of Economic Geography* 5 (2): 119–53.

Glaeser, Edward L, and Albert Saiz. 2003. "The Rise of the Skilled City." NBER Working Paper 10191.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy* 118 (3): 536–98.

Gross, Daniel P., and Bhaven N. Sampat. 2023. "America, Jump-Started: World War II R&D and the Takeoff of the US Innovation System." *American Economic Review* 113 (12): 3323–56.

Gruber, Jonathan, and Simon Johnson. 2019. *Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream*. PublicAffairs.

Guzman, Jorge, Fiona Murray, Scott Stern, and Heidi Williams. 2024. "Accelerating Innovation Ecosystems: The Promise and Challenges of Regional Innovation Engines." *Entrepreneurship and Innovation Policy and the Economy* 3: 9–75.

Hausman, Naomi. 2022. "University Innovation and Local Economic Growth." *Review of Economics and Statistics* 104 (4): 718–35.

Helmers, Christian, and Henry G. Overman. 2017. "My Precious! The Location and Diffusion of Scientific Research: Evidence from the Synchrotron Diamond Light Source." *Economic Journal* 127 (604): 2006–40.

Holman, Mary A., and Ronald M. Konkel. 1968. "Manned Space Flight and Employment." *Monthly Labor Review* 91: 30–36.

Jaffe, Adam B. 1989. "Real Effects of Academic Research." *American Economic Review* 79 (5): 957–70.

Johnson, Rucker C., and C. Kirabo Jackson. 2019. "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending." *American Economic Journal: Economic Policy* 11 (4): 310–49.

Kantor, Shawn, and Alexander Whalley. 2014. "Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks." *Review of Economics and Statistics* 96 (1): 171–88.

Kantor, Shawn, and Alexander Whalley. 2019. "Research Proximity and Productivity: Long-Term Evidence from Agriculture." *Journal of Political Economy* 127 (2): 819–54.

Kantor, Shawn, and Alexander Whalley. 2025. "Moonshot: Public R&D and Growth." *American Economic Review* 115 (9): 2891–925.

Kerr, William R., and Frederic Robert-Nicoud. 2020. "Tech Clusters." *Journal of Economic Perspectives* 34 (3): 50–76.

Lee, Jongkwan. 2019. "The Local Economic Impact of a Large Research University: Evidence from UC Merced." *Economic Inquiry* 57 (1): 316–32.

Lerner, Josh. 2009. *Boulevard of Broken Dreams: Why Public Efforts to Boost Entrepreneurship and Venture Capital Have Failed—and What to Do About It*. Princeton University Press.

Lerner, Josh, Henry Manley, Carolyn Stein, and Heidi Williams. Forthcoming. "The Wandering Scholars: Understanding the Heterogeneity of University Commercialization." *Econometrica*.

Liu, Sitian, and Yichen Su. 2026. "The Geography of Research: The Trade-Off between Knowledge Production and Access." Preprint, arXiv. https://doi.org/10.48550/arXiv.2608.15981.

Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. 2023. *IPUMS National Historical Geographic Information System: Version 18.0* [dataset]. IPUMS Center for Data Integration. http://doi.org/10.18128/D050.V18.0 (accessed June 28, 2024).

Marco, Alan C., Michael Carley, Steven Jackson, and Amanda Myers. 2015. "The USPTO Historical Patent Data Files: Two Centuries of Innovation." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.2849514.

Markoff, John. 2022. "The Roots of Silicon Valley, Part 1: Founders, Legend, Legacy." *EE Times*. https://www.eetimes.com/the-roots-of-silicon-valley-part-1-founders-legend-legacy/ (accessed June 28, 2024).

Moretti, Enrico. 2004a. "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics* 121 (1–2): 175–212.

Moretti, Enrico. 2004b. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review* 94 (3): 656–90.

Moretti, Enrico. 2012. *The New Geography of Jobs*. Houghton Mifflin Harcourt.

Myers, Kyle R., and Lauren Lanahan. 2022. "Estimating Spillovers from Publicly Funded R&D: Evidence from the US Department of Energy." *American Economic Review* 112 (7): 2383–423.

Nelson, Richard R., ed. 1993. *National Innovation Systems: A Comparative Analysis*. Oxford University Press.

NPR. 2008. "Timeline: A Look at Bill Gates' Life." https://legacy.npr.org/news/graphics/2008/june/bill_gates/gates_timeline_04.html (accessed June 10, 2024).

PBS. 1999. "William Shockley Moves to Mountain View." https://www.pbs.org/transistor/background1/events/shockleymove.html (accessed June 25, 2024).

Petralia, Sergio, Pierre-Alexandre Balland, and David L. Rigby. 2016. "Unveiling the Geography of Historical Patents in the United States from 1836 to 1975." *Scientific Data* 3: 160074.

Ramey, Valerie A. 2019. "Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?" *Journal of Economic Perspectives* 33 (2): 89–114.

Saxenian, AnnaLee. 1994. *Regional Advantage: Culture and Competition in Silicon Valley*. Harvard University Press.

University of California, Merced. 2014. "Campus's Direct Investment in Regional Economy Exceeds $1.1 Billion." University of California, Merced Newsroom, October 13. https://news.ucmerced.edu/news/2014/campus%E2%80%99s-direct-investment-regional-economy-exceeds-11-billion.

US Census Bureau. 2024. *County Business Patterns: 2019*. https://www.census.gov/data/datasets/2019/econ/cbp/2019-cbp.html (accessed June 28, 2024).

US Patent and Trademark Office. 2024. *Data Download Tables*. US Patent and Trademark Office Open Data Portal. https://patentsview.org/download/data-download-tables (accessed June 28, 2024).

Valero, Anna, and John Van Reenen. 2019. "The Economic Impact of Universities: Evidence from across the Globe." *Economics of Education Review* 68: 53–67.

Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review* 88 (1): 290–306.

Journal of Economic Perspectives—Volume 40, Number 4—Fall 2026—Pages 53–72

Trust Violations in Research Universities

William G. Howell

In all sorts of ways, the research operations of colleges and universities rely upon public trust. Sabbaticals and fellowships routinely offer support and protections for open-ended inquiry that, the public is told, will eventually yield insights and discoveries that serve the common good. The institution of tenure is predicated on a belief that professors can be trusted to continue to work hard and work well for the duration of their careers. Arguments against various forms of professional monitoring and oversight rely upon an assurance that we, professors, know what we're doing, that we are trustworthy, that the world is well-served by investing in us—and then leaving us alone.

So universities need trust. But for years, this trust has fallen. A bounty of survey data shows that trust in universities has declined substantially over the last two decades, and more quickly than other public institutions.

Many factors have contributed to trust's downfall. Steadily rising tuition costs coupled with diminishing employment prospects for university graduates, rising populism, and enduring strains of anti-intellectualism all are contributing factors. Both at home and abroad, aspiring autocrats wishing to consolidate their power have put research universities in their sights. In ways large and small, universities have become entangled in larger political fights and social transformations that spin up all sorts of opponents and detractors.

But in this maelstrom, which features ample lies, distortions, and arguments in bad faith, universities are not mere victims. As others have recognized (Balafoutas

■ *William G. Howell is Dean of the School of Government and Policy and Professor of Political Science, Johns Hopkins University, Baltimore, Maryland. His email address is whowell5@jh.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251484.

et al. 2025), a series of replicability crises has done real damage to the credibility of certain disciplines, and so has the scientific community's occasional tendency to exaggerate the evidentiary basis for controversial policies, including stay-at-home orders during the Covid pandemic (Macedo and Lee 2025).

Two additional factors threaten to further erode public trust in the academic research enterprise: the blurred distinction between scholarship and advocacy and a professoriate that supports an increasingly narrow range of ideological viewpoints. They are the focus of this essay. Research activism and ideological compression are not endemic everywhere, of course. Universities are a diverse lot, and each is a complex organization. But across higher education, these violations have occurred enough times in enough places to cause real damage.

If trust is to be restored, it's important that universities not merely admit missteps and mistakes—begrudgingly, and only when pressed. Without prompting or threat, they should acknowledge failures. Universities must see themselves fully. And they must change their ways, lest others, who care not for their welfare and believe not in their mission, force the issue.

Research Universities Need Trust

It is an audacious proposition that universities put before their various publics—one that appears in two parts. Invest in ever-increasing amounts, universities first insist, so that they might grow in size, stature, and influence; and so, the public has. Hundreds of billions of dollars from philanthropies and governments pour into university coffers each year. State and local governments supported higher education to the tune of $149 billion in 2025 (State Higher Education Executive Officers Association 2025); federal support for university-based research and development was $60 billion (National Science Board 2025), while federal Pell grants were another $38 billion (Humphrey 2025). Voluntary giving to higher education reached $78 billion in 2025 (Council for Advancement and Support of Education 2026). Moreover, because universities are nonprofit institutions, the donors receive tax breaks and universities are often at least partially exempted, under a variety of state and local laws, from sales and property taxes.

Having received donations, investments, and benefits that surpass the tuition payments from households of current students, universities then insist that accountability come only from within—and it is here that their proposition becomes audacious. After writing their checks, universities argue, governments and donors should stand down. Do not tell them what can be taught; who can be hired, promoted, or fired; what research questions are asked or ignored; what conclusions are drawn; and how these conclusions are communicated, or to whom. All of these are matters exclusively for internal consideration and review. To even hint that a funder, legislator, or parent might dictate agendas from afar is to violate the research university's most sacrosanct commitment to academic freedom.

Academic freedom consists of more than norms and expectations. It is built into the business model of higher education. It appears in explicit provisions barring interference by funders in the contracts, grants, and gifts that universities negotiate. It is fortified by the institutions of tenure and faculty governance. It is buoyed by unrestricted fellowships and research sabbaticals that are regularly granted to faculty. It manifests in both the foreground and background of university mission statements. It is baked into the very identity of nearly everyone involved in the research enterprise.

No other public institution has brokered such an exquisite bargain, one that confers both largesse and freedom. How have universities managed to do so? In part, it is because of the exceptional value propositions that they offer (Geiger 2004; Cole 2009). To governments, universities highlight the patents, cures, discoveries, and diffusion of technological progress that come from within their ranks. To philanthropists, universities point to the advances made in combating all manner of real-world problems. To parents, universities insist that their children will graduate gainfully employed, well compensated, networked, stimulated, and civilized. These accomplishments are real. Universities have a great deal of which to be proud.

Past achievements and triumphs, however, don't entirely settle the question. After all, the promises that universities put before their publics, particularly as they relate to the conduct of academic research, are fraught with ambiguity and uncertainty. Presumptively, it remains unclear what investments in university research will turn up—particularly in domains that do not rely upon competitive grants to fund basic operations. In the humanities and social sciences, most notably, there are no guarantees that any particular finding will be discovered, that any specific puzzle will be resolved, or that any particular question will be fully answered. As researchers struggle to make progress, their discoveries assuredly benefit from foresight and planning, but also from happenstance, accident, serendipity, and even luck. As the aphorism goes, "If we knew what it was we were doing, it would not be called research, would it?"

Academic research rarely progresses along neat, narrow, and linear pathways. Rather, all sorts of spillovers characterize the enterprise, as inquiry aimed at one problem often generates insights that bear on entirely different ones. As Azoulay, Zivin, Li, and Sampat (2019) show, for instance, research funded by the National Institutes of Health to study one disease is just as likely to produce drug patents for others. In such ways, the learning and benefits of academic research extend well beyond any investigator's original intent.

Because of its uncertainty, vast domains of the academic research enterprise are not devoted to delivering individual knowledge claims, one at a time, on time and on budget, each prenegotiated in advance. Rather, these domains cultivate research ecosystems in which truths and knowledge are pursued flexibly and improvisationally. They deliberately and conscientiously invest in preconditions that enhance the probability that a researcher might stumble across some new fact or pattern—one that no one imagined in advance, but one that redefines ways of thinking, understanding, and even living after it has occurred. As Joan Wallach Scott

(2019, p. 13) puts it, the production of knowledge in higher education follows an "open, unending process by a faculty 'sifting and winnowing' its truth claims." This process demands a "certain discipline and rigor," Scott reminds us, but it cannot be reduced to a set of fixed instructions. Seeking to prescribe the pathways and objects of truth-seeking that universities support risks diminishing the entire enterprise. Researchers, instead, must be afforded the space, freedom, time, and protection to pursue hunches, to improvise, and to experiment.

While open-ended, academic research gains traction by adopting what Robert Merton (1973) famously called "organized skepticism": a disciplined commitment to questioning all claims, including those held dearly, and to suspending external or personal interests in favor of impartial evaluation. Merton emphasized that this norm, together with others like communalism and disinterestedness, legitimizes and propels the research enterprise. The result, Merton tells us, is not a straightforward march toward truth but a collective, iterative endeavor over which no single individual or entity is in charge.

Universities, of course, are not the only institutions that undertake research for public purposes. Contractors, think tanks, and the government itself all perform the task. Usually, though, they do so in ways that are considerably more circumscribed than what either Scott (2019) or Merton (1973) describe—evaluating a program, surveying a population, investigating a security threat, and conducting numerous other studies that are contracted and undertaken for relatively well-defined, instrumental purposes. It is no coincidence, moreover, that these alternative organizations do not offer the kinds of job protections or academic freedoms that support academic research. Only universities promise lifetime employment to certain favored employees and considerable freedom to ask questions, study subjects, and teach materials of their own choosing. No other public institution sets its research interests and ambitions so broadly.

Within universities, meanwhile, one can certainly find various forms of oversight and obligations—particularly in soft-money environments that depend upon outside philanthropic support and government contracts. There, faculty seeking outside support must make their case for a research program's importance and then promise a set of deliverables under a well-defined timetable. Notice, though, that these obligations come from outside universities and appear quite distinct from an internal environment in which academic research, particularly within the humanities and social sciences, is remarkably devoid of such encumbrances. Hiring and promotion decisions, of course, focus on specific elements of a scholar's record of accomplishments. But in between, when the bulk of research at universities is actually conducted, faculty are free to pursue questions of their choosing through means they select at a pace that they alone deem appropriate in the service of formulating answers to their liking—not perfectly, of course, but to an extent unrivaled elsewhere. It is unfathomable that a department chair, divisional dean, or provost would hand down strict instructions intended to fix a faculty member's research agenda. In what other environment does the leadership of an institution exercise so little control over the employees it pays?

The special case that universities make for investment and autonomy, therefore, crucially depends upon trust—trust that professors know enough of what they're doing to steer their efforts in directions that have a higher probability of social payoffs; trust that they will pursue the truth wherever it might lead; and trust that the resulting research eventually will turn up lessons, discoveries, inventions, and insights that may not have been prespecified or contractually stipulated, but that nonetheless enhance human knowledge and, by extension, human flourishing. Ambiguity and uncertainty can remain baked into the academic research enterprise only so long as publics are willing to trust the process and resist the impulse to interfere. Arguments by universities against various forms of professional monitoring and oversight rely upon an assurance that we, professors, know what we're doing, that we are trustworthy, and that the world is well served by investing in us and then leaving us alone.

Trust us, universities implore, and good things will follow.

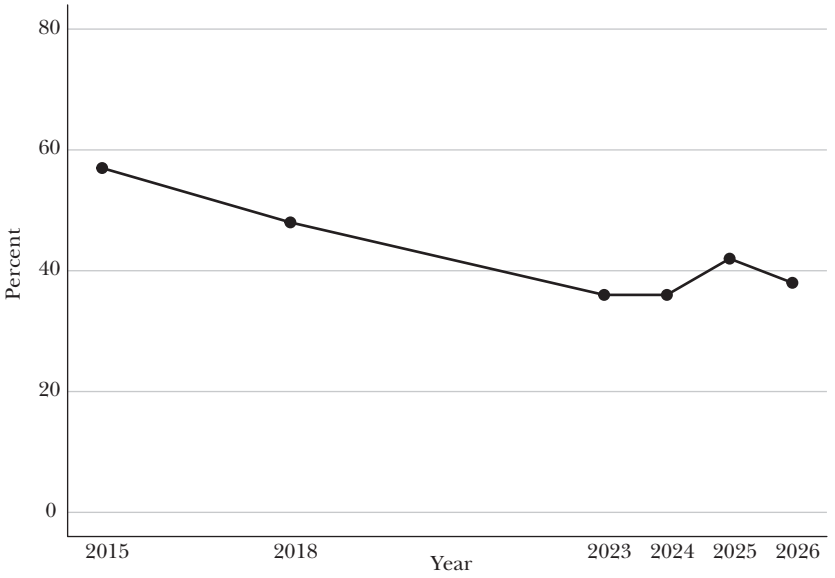
The Decline of Trust

For much of the modern era, the American public did trust universities to undertake the research enterprise. And many good things did follow: life-saving medical breakthroughs; scientific discoveries about the universe's origins, plate tectonics, and the workings of the human brain; new technologies in agriculture, energy, transportation, communication; facts and theories about our individual and social existences; and on and on (Rosenberg and Nelson 1994; Mowery, Nelson, Sampat, and Ziedonis 2001; Slaughter and Leslie 1997; Cohen, Nelson, and Walsh 2002; Azoulay, Zivin, Li, and Sampat 2019). Notwithstanding these achievements, the compact between universities and the public is more fragile today than it has been in decades. Measured in any number of ways, trust in higher education has fallen.

Starting in 2015, the Gallup Poll has occasionally asked nationally representative samples of adults living in the United States about their confidence in higher education (Jones 2026). As shown in Figure 1, in 2015, 57 percent of respondents claimed to have either "a great deal" or "quite a lot" of confidence in higher education. Eleven years later, that number registered 19 percentage points lower.

The drop was not confined to any particular demographic group, whether measured by education, income, sex, or geography. Broad swaths of the American public now express lower levels of confidence in colleges and universities than they did just a decade prior. However, the Gallup numbers did reveal some partisan differences. Both the lowest level of confidence and the largest recorded drop from 2015 to 2026, by far, appear among Republicans. In 2015, 56 percent of Republicans expressed confidence in higher education. About a decade later, only 23 percent did so. Nonetheless, the confidence levels of other groups fell as well. Independents and Democrats also show noticeable declines. In 2015, 48 and 68 percent of Independents and Democrats expressed confidence in higher education, respectively; in 2026, 39 and 50 percent did so.

Figure 1
Declining Public Confidence in Higher Education



Source: Gallup.
Note: Percent of respondents reporting "quite a lot" or "a great deal" of confidence in colleges and universities, 2015–2026. Polling only conducted during years shown.

Of course, declining public trust in US institutions is not confined to universities. In the last several decades, numerous organizations and institutions have suffered a loss of public trust. Confidence levels in the military, the police, public schools, the business sector, the media, political branches of government, and the medical establishment have all declined. Confidence in Congress, never particularly high, is now at historic lows. Rather extraordinarily, however, over the last decade universities experienced by far the largest drop in public confidence. Between 2015 and 2026, the decline registered for universities was nearly twice that observed for any other institution that Gallup surveyed.[1]

These trends do not appear to be an artifact of any single sampling procedure or question wording. Alternative polling reinforces these same patterns. The 2024 FIRE/NORC survey found just 28 percent of Americans report "a great deal" or "quite a lot" of confidence in universities, while 30 percent now say they have "very little or none at all" (as reported by Honeycutt 2024). When the Gallup Poll probed for reasons for skepticism about higher education, a plurality of Americans cite concerns about the political agendas of universities (31 percent), with

[1] Gallup, "Confidence in Institutions," https://news.gallup.com/poll/1597/confidence-institutions.aspx?. See also Druckman et al. (2026), who observe a larger drop in public trust in the Supreme Court than other public institutions, as measured by the National Election Studies and General Social Surveys.

concerns about cost and free speech trailing just behind (Jones 2026). Research by the Association of American Universities sheds further light on the matter: While 63 percent of Americans trust research universities to pursue life-saving science, only 29 percent believe they teach without political bias.[2]

The downward trend in public trust likely extends back beyond the last decade. Longer-running datasets suggest that trust started declining decades ago. The General Social Survey (2025) finds that confidence in all educational institutions fell from well over 40 percent in the early 1970s to the low 20s in recent years. Other Harris polling data align with these findings (Brady and Kent 2022).

Assuredly, no single survey finding is dispositive. The quality of the surveys and the representativeness of the samples all vary, as does the focus of the questions themselves. In some instances, respondents are asked about trust, in others about confidence, and in others something different. The meaning of these questions, too, can be unclear. It is hard to discern what, exactly, respondents have in mind when they say they "trust" universities. Trust universities to do what? Trust relative to what?

Still, amidst all this variability and ambiguity, a singular point comes through: The public's trust—confidence, support, or faith, call it what you will—in universities is in steep decline. While the overall environment reflects a generalized weariness with traditional civic pillars, universities and colleges have become flashpoints in the country's trust deficit. Americans are less trusting of nearly all civic institutions, but none has experienced such a rapid and widespread loss of confidence.

Trust Violations

Universities are not mere bystanders to trust's decline. They have acted in ways that certainly justify, and possibly have helped to cause, recent trends: first, by pursuing research as activism; and second, by allowing the range of ideological perspectives in departments and divisions to narrow. Neither of these violations is incidental to the research enterprise. Rather, both abrogate the research university's central obligation to seek, discover, document, and communicate knowledge and truth, which must be understood as the overriding mission of academic research and justification for its funding and attendant privileges and freedoms.

Trust Violation 1: Research as Activism

The first violation of the public trust by research universities takes the form of activism. Given the ideological composition of the faculty today, a topic discussed at

[2]Goldstein, Kenneth, Hannah Ritz, Graham Andrews, and Amanda Shaffer. 2024. "Public Opinion Update." *Association of American Universities*. https://www.aau.edu/uploads/files/general/PC-Final-Deck-October-2024.pdf. See also: Reuters, Ipsos. 2025. "Reuters/Ipsos Large Sample Poll February 2025." https://www.ipsos.com/en-us/reutersipsos-large-sample-poll-february-2025; Nguyen, Sophie, Rachel Fishman, and Olivia Cheche. 2024. *Varying Degrees 2024*. New America. https://www.newamerica.org/insights/varying-degrees-2024/.

length below, this activism now comes predominantly from the Left. But its source is immaterial, at least when it comes to matters involving trust. Whenever professors understand their research obligations as something other than the pursuit of truth and advancement of knowledge, their scholarship threatens the pact between the public and higher education.

In some instances, activism takes hold as professors become disabused of the very idea of truth and instead see power relations everywhere. In other instances, looking out upon a world filled with failures, harms, and injustices, professors feel compelled to treat their scholarship as an instrument for change. Some professors even lose sight of the larger mission of universities, and their politics take hold of everything they do. But no matter the reason or rationale, the outcome is the same: When professors trade in the work of knowledge creation and dissemination for political advocacy over which they hold no special authority, they violate the public trust.

To be clear, the trouble does not come from professors who engage in public protest, who donate to politicians, who harbor strong and definite opinions about the next election, who take a leave of absence in order to serve in one presidential administration or another, or who are now or have ever been members of the Communist Party. When it comes to public trust in the research enterprise in higher education, all of that is (or should be) immaterial.

Trouble comes, instead, through other channels: when professors rotate back into academia and embed their research within a political project, no matter how modest or mainstream; when they censor certain findings because of the problems they'll generate for their chosen cause; when they refuse to ask certain questions for fear of upsetting political allies or pleasing enemies; when they evaluate the benefits of a policy intervention without ever considering its costs, or costs without considering benefits, lest they dignify the objections of political opponents; when they go out of their way to valorize preferred groups while reflexively ascribing malign intent and moral depravity to those in positions of power and privilege; when they selectively recognize the agency, culpability, or mistakes of a certain population according to how its members figure in a larger political project; when they profess that their research serves—not distantly or abstractly, but quite directly—an activist agenda to eliminate injustices, right wrongs, recover heritages, glorify gods, or uphold some particular understanding of the public good.

Khaitan (2022, p. 547) refers to this pattern of actions, orientations, and dispositions as "scholactivism." As he defines it, "Whereas truth-seeking and knowledge dissemination are constitutive of the role of a scholar, scholarly-activism-driven research is distinguished by the existence of a motivation to directly pursue specific material outcomes (that is, outcomes that are more than merely discursive) through one's scholarship." For the scholactivist, the search for truth and the creation of knowledge no longer anchor and guide research. Rather, they are subordinated to, and sometimes even replaced by, political imperatives.

The problem with these actions and orientations is not that they may lead professors, in their own time, to articulate politically controversial positions. Nor do problems necessarily arise when the contents of academic research become fodder

for various political campaigns. Such are professional hazards even when research is completely aligned with truth-seeking and knowledge creation. The problem, instead, is that these actions and orientations diminish the credibility of universities by running afoul of the very arguments that are routinely made for additional investments in and protections for academic research.

Consider the strong case often made for the complementary institutions of tenure and academic freedom (for variants, see Reichman 2021; Lackey 2018; Hertzog 2017). It runs basically as follows. Having demonstrated a high level of excellence in knowledge and achievement, professors should be granted lifetime appointment not in honor of their early work, but in order to free them up to pursue new knowledge and discoveries without fear of political interference or retaliation. When conducting their research, these professors should be allowed to seek truth and advance knowledge wherever it might lead and without concern for how it will accord with the political objectives of a donor, dean, president, or politician. Tenure, its advocates rightly argue, increases the chance that professors will assume risks in their scholarship, that they will undertake projects that may or may not pay off, and that they will see the objects of their study more fully and more deeply, given assurances that they will remain gainfully employed regardless of what their studies turn up.

Against this backdrop, the offense of scholarly activism is plain to see. Lifetime employment and academic freedom are meant to offer protection from other people's politics so that faculty might discover new truths and create new knowledge—so that they might devote their research to what Veblen (2015[1918], p. 51) refers to as "disinterested learning, given over to the pursuit of intellectual inquiry." That faculty are *dis*interested, of course, does not mean they are *un*interested. Scholars care greatly about the objects of discovery. But their research must be faithful to these objects themselves and not the political purposes to which they might be put. As Veblen (p. 62–63) explains, "the spirit of disinterested inquiry must have free play in these seminaries of the higher learning, without afterthought as to the practical or utilitarian consequences which this free inquiry may conceivably have for the professional training or the social, civil or religious temper of the students or the rest of the community." Good things may come from disinterested research, to be sure, but such outcomes appear as afterthoughts and welcomed rewards, not guiding lights.

This ideal of disinterested learning was central not only to Veblen's conception of the university, but also to the foundational arguments for academic freedom offered long ago by the American Association of University Professors (AAUP). The AAUP's original "Declaration of Principles on Academic Freedom and Academic Tenure" (henceforth, the Declaration) was born from concern that professors would be unable to build knowledge and pursue truth if their inquiries were subject to political, financial, or institutional pressures.

The Declaration, first drafted by a joint committee of the nation's economics, political science, and sociology associations in 1913, and then formally adopted by the newly established the American Association of University Professors in 1915,

defined academic freedom as encompassing three essential rights: the freedom of inquiry and research; the freedom of teaching; and the freedom of extramural utterance, which allowed faculty, in their own time, to be as political as they choose. However, just as important as these rights, the authors insisted, were the professional obligations that accompanied them. Faculty must approach their work in a "scholar's spirit," committed to "competent and patient and sincere inquiry." Sincerity, here, does not merely call for honesty or personal conviction. Rather, it was a methodological commitment—a disciplined effort to reach conclusions through rigorous, unbiased inquiry, free from pecuniary inducement or ideological obligation. Professors were to resist any motive "other than their own scientific conscience and a desire for the respect of their fellow experts." To compromise this principle was, in the words of the Declaration, to "corrupt" the university's teaching and research obligations, to "vitiate" its appropriate influence on public opinion, and to deprive society of the "peculiar and necessary service" that only the professional scholar could render.

The drafters of the Declaration did not deny that scholarship—especially in the social sciences—would have political implications. Nor did they expect professors to be politically disengaged. What they feared, and what they sought to guard against, was the interference of political commitments with scholarly judgment. The whole purpose of tenure and academic freedom was to ensure that politics would not contaminate the scholar's pursuit of knowledge and truth. The leaders of a university, they insisted, have "no moral right to bind the reason or the conscience of any professor." But the conscience of the professor, acting in the role of a professor, was to be anchored by the disinterested pursuit of knowledge.

The research activist reneges on this bargain. While enjoying all the resources, job protections, and freedoms that offer a shield from other people's politics, the activist deploys this bounty in the service of their own personal politics. It's an extraordinary bait and switch, and critics are quite right to call foul. What makes this reversal especially damaging is that it undermines the very justification for tenure and academic freedom in a politically heterogenous society. Universities are granted their extraordinary privileges with the understanding that they will hold themselves to a higher standard of scholarly integrity. When faculty abandon that standard and use their position to advance their own political causes rather than cultivate knowledge, they forfeit public trust and invite demands for external oversight, regulation, and reform. Activism distorts the purpose of research just as it threatens the very conditions that make independent scholarship possible in the first place.

How often does this happen? It's hard to say. Some faculty plainly admit their activist intentions, and others openly call for a fundamental reorientation of disciplines and fields of study around "political projects" (for example, Gupta and Stoolman 2022). Such views and intentions, however, are not routinely and openly admitted (Wright, Hecht, and Saxe 2025). Instead, they are sublimated. Rather than being stated forthrightly, activism percolates in the subterranean layers of their research—in the questions asked, the emphases placed, the framing offered, the implications recognized. This activism further shape the kinds of quiet mentorship

that is offered to graduate students about the value of different kinds of research, the notions of "reasonableness" that shape scholarly exchange, understandings about what makes a research agenda worthy of study, what makes a research contribution a big deal or mere trifling. This activism also pervades the peer review process, wherein certain politically attractive findings are put on the fast-track to publication in a discipline's very best journals, and politically unpopular findings struggle to find acceptance, no matter their scientific merit.

Impressionistically, the problem of activism appears to be more pronounced in the humanities and social sciences than in the hard sciences. In a recent essay, Clune (2024) reflects on the state of affairs in English departments. He observes: "Professors began to see the traditional values and methods of their fields—such as the careful weighing of evidence and the commitment to shared standards of reasoned argument—as complicit in histories of oppression. As a result, many professors and fields began to reframe their work as a kind of political activism." The historian David Blight observes that "there are tendencies these days among younger historians, but not just younger historians, to sometimes put advocacy ahead of scholarship" (as quoted by Green 2023). Reflecting on recent trends within sociology, Savolainen (2025) observes that "the movement for 'public sociology' has encouraged researchers to align their work with social justice activism of a decidedly anticapitalist persuasion." Similar comments could be made about other fields of study.

Moving beyond anecdotes and testimonials to quantify these trends systematically is notoriously difficult (but for one noteworthy attempt, see Rubin 2025). Taking the additional step of parsing the particular influence of activism—as distinct from prevailing intellectual currents, disciplinary fashions, professional incentives, or any number of other factors affecting the research enterprise—is nearly impossible. Without a clear sense of the problem, then, claims and counterclaims are bandied about. Critics of universities are quick to say that activism is pervasive, defenders insist that it is entirely exceptional, while the truth of the matter remains unclear.

Lest you think it doesn't exist at all, though, the promotion of an activist agenda is sometimes openly declared. *Nature Human Behaviour* is a well-regarded journal in the social, biological, and health sciences. In 2022, the journal published an editorial titled "Science Must Respect the Dignity and Rights of All Humans." The journal then introduced ethics guidance that would allow editors to reject, modify, or even retract already-published research if they deemed it could harm marginalized groups, regardless of the scientific quality or accuracy of the work. The policy specifically stated that editors could take these actions "regardless of content type (research, review or opinion) and, for research, regardless of whether a research project was reviewed and approved by an appropriate institutional ethics committee."

Some academic departments also openly declare their intention to promote an activist political agenda in all of their work, including research. Examples aren't hard to find in their mission statements and on their homepages. For American University's sociology department, "Our common purpose is education and research for

social justice." Scholarship at the Department of English at Michigan State University is "explicitly committed to advancing equity and justice and questioning established power structures." Faculty in Clark University's English Department "commit to using our own power as scholars and teachers to resist and dismantle systems of oppression" and to advance the causes of "anti-racism, disability justice, feminist practice, anti-classism, and the affirmation of LGBTQ+ identity and experience."[3]

Outside of formal academic departments, some professional associations are assuming a distinctly activist posture. The American Anthropological Association, for instance, announced in 2020 that its "research, scholarship, and practice" should be placed "in service of dismantling institutions of colonization and helping to redress histories of oppression and exploitation."[4] In 2024, the theme of the Annual Meetings of the American Sociological Association was "Intersectional Solidarities: Building Communities of Hope, Justice, and Joy," with the intended purpose of promoting "sociology as a form of liberatory praxis: an effort to not only understand structural inequities, but to intervene in socio-political struggles," as quoted in Lee (2026).

What happens when a scholar uncovers findings that are inconvenient for, or even challenge, a political agenda devoted to decolonization, liberation, or social justice? What if they know that a professor who favors the stated mission is a journal editor, or a potential referee for a paper, or part of a hiring committee? Do they conjure reasons to downplay uncomfortable findings? Do they hide them from public view? Or do they frame their questions and pursue lines of inquiry that minimize the chances of their discovery in the first place?

Many faculty members, I hasten to say, bridle at all those options. Most, I'm inclined to believe, stay true to Veblen's mandate. But not all. And when they deviate, as activist mission statements counsel, researchers violate the trust upon which their tenure and academic freedom depend.

Trust Violation 2: Ideological Compression

It's unavoidable, of course, that values, beliefs, and politics seep into scholarship—in the topics studied, the standards of evidence held for different knowledge claims, the meanings assigned to different fact patterns, and even the findings uncovered (Borjas and Breznau 2024; Jelveh, Kogut, and Naidu 2024; Manzi 2026). While research activism may be relatively unusual in higher education, social and political biases are not. And while their existence is unavoidable—we are human, after all—the research university's ability to advance truth and build knowledge is compromised when these biases are not identified, challenged, and addressed.

[3] For links to these statements, see American University, Department of Sociology, "Home," accessed August 5, 2025, https://www.american.edu/cas/sociology/. Michigan State University Department of English, "Equity & Justice Mission," accessed August 5, 2025, https://english.msu.edu/equity-justice-mission/. Clark University, "English Department Statement," accessed August 18, 2025, https://www.clarku.edu/departments/english/english-department-statement/.

[4] As quoted in Rose Horowitch, "The Era of DEI for Conservatives Has Begun." *The Atlantic*, May 27, 2025.

What is to be done? Much of the onus rightly falls upon individual scholars. Researchers must do their level best to spot their biases and mitigate their influence. When biases do more than just shape the research questions asked and instead interfere with the effort to formulate well-founded answers, scholars should try to eradicate them. Scholars should make a point of asking themselves questions that help to illuminate them. (For example, would I think differently about the value of conducting this study had the findings come out exactly opposite?) They should attempt to stand outside of themselves and view the objects of their study from a position of impartiality—inhabiting what Thomas Nagel (1986) refers to as the "view from nowhere," wherein every insight is objective, every observation true.

Scholars have obligations here. And universities must insist that they try to fulfill them. Still, for most, these efforts inevitably falter. Biases are too deeply rooted in people's identities and habits of mind to leave their research untouched. They shape not only their preferences but their perceptions—what they notice, what they understand, what they take for granted. Scholars are not transparent to themselves.

So auxiliary precautions are needed. While calling upon their faculty to rise above their individual biases, as best they are able, research universities also must acknowledge that scholars will never reach a place of complete transcendence, and that they require communities and practices that treat this condition not as a moral failing, but as a methodological vulnerability. To do so, universities must cultivate cultures and communities in which faculty seek out disagreement rather than avoid it, submit their claims to the scrutiny of others who are unlikely to reflexively agree. For contentions to become knowledge, for idle conjectures to be distinguished from established truth, they must survive contact with dissent. It's not the absence of biases that sustains the integrity of academic research, but the presence of interlocutors who compel their confrontation.

Here we see the urgent need for intellectual pluralism in higher education. Pluralism, of course, can be understood as a value unto itself, a point of pride for universities seeking to represent and serve the broader communities in which they operate. But pluralism also has instrumental value for the university research enterprise, ensuring "vigorous skeptical vetting of scientific claims by a critical mass of doubters" (Abbot et al. 2023, p. 3). As long as a diverse assembly of researchers stand watch over one another, biases can be made to check biases—much as ambitions are made to check ambitions in the Madisonian system of the US government—and the work of truth-seeking is enhanced. Researchers should engage others who think differently across ideological, methodological, and disciplinary lines—not as an act of charity or tolerance, but as a prerequisite for scholarly progress. While maintaining strict standards of excellence, universities must nurture and then unleash a kind of "purposeful pluralism" that forces scholars to not just recognize and tolerate difference, but to put such difference to good use (Daniels 2021). By deliberately opening their research to critique from people who do not share their assumptions, faculty improve their chances of getting things right and reduce the risk of mistaking ideological priors for insight.

When conducting research, scholars mustn't be left alone with their biases. Worse still, they mustn't have their biases merely reinforced by likeminded company. Instead, they must submit to rigorous scrutiny from diverse perspectives, what Helen Longino (1990) calls "transformative criticism." This collective critical engagement requires a shared commitment to openness and intellectual humility, allowing the research enterprise to self-correct over time. Longino argues that without this trust in the community of scholars—who are free to investigate, challenge, and revise claims—research risks stagnation or ideological capture. The material support and autonomy granted to academics is not a license for faculty to surround themselves with people who are predisposed to affirm and ingratiate, but to build pluralistic communities in which dissent is common, biases are caught, diagnosed, and mitigated, and the pursuit of truth advances.

Have universities built and sustained these communities? Perhaps once, but not now. The situation appears to be growing worse.

Evidence of a compression of ideological perspectives in higher education, particularly in the humanities and social sciences, is indisputable and overwhelming. The professoriate does not merely sample disproportionately from one end of the ideological spectrum. It includes, nearly exclusively, people from the near, middle, and far Left. Centrist and independent faculty, for the most part, are significantly outnumbered and, fearing backlash and feeling cowed, do their best to blend into campus crowds. Conservative faculty, meanwhile, are endangered species who, to avoid detection, keep their heads bowed and their opinions quiet (Shields and Dunn 2016).

A wide variety of empirical studies point to a common conclusion: Across a range of institutions and disciplines, a pronounced ideological imbalance characterizes university faculty today. Analyzing over 12,000 professors at public and private universities, Langbert and Stevens (2020) found a Democrat-to-Republican party registration ratio of more than 8 to 1. This imbalance persists across academic disciplines. Anthropology shows the most skewed ratio at 42 to 1, while even economics—the most politically balanced discipline—registered a 3 to 1 ratio. Among faculty who make political donations, the overall disparity was even more stark, with a partisan ratio of 95 to 1.

These imbalances are not confined to public universities. A 2018 study of elite liberal arts colleges found that 39 percent of institutions employed either none or a miniscule number of registered Republican faculty members (Langbert 2018). Among those that did, 78 percent of departments did not have even one Republican faculty member.

Longitudinal data document an ideological tilt that is not only enduring but intensifying. The long-running faculty survey carried out by the Higher Education Research Institute, for example, shows a steady leftward drift over the past three decades, with the share of professors identifying as "liberal" or "far left" consistently outpacing national trends. In its 2016 survey, 60 percent of faculty placed themselves on the Left, compared to just 12 percent on the Right—a gap that has widened since the 1980s. A qualitatively similar finding from Gross

and Simmons (2007) suggested 44 percent of professors identified as liberal, as compared to conservatives at 9 percent. According to a more recent report by Nathan Honeycutt (2025), liberal faculty outnumber conservative faculty nationwide by a margin of more than 7 to 1; and the number of "far Left" and "very liberal" faculty members exceeds the total number of faculty sitting anywhere on the Right of the political spectrum.

Also using campaign contributions, David Primo (2026) tracks the ideological distributions of faculty at 55 research universities and compares them to legislators in the US Congress. His findings again reveal a stark ideological skew: "Using the same measure of ideology for legislators and faculty contributors, Representative Alexandria Ocasio-Cortez and Senators Bernie Sanders and Elizabeth Warren would rank as the furthest-left members of their respective chambers in 2024, yet would appear only slightly left of center in a legislature composed of faculty contributors."

These imbalances extend into college administrations and professional schools as well. A 2018 NORC survey of over 900 student facing administrators reported that liberals outnumbered conservatives by roughly 12 to 1 (Abrams and Grigorian 2018). At top-ranked public policy schools, a 2024 study found that demonstrably liberal faculty outnumbered demonstrably conservative faculty by 7 to 1 (Hess and Fletcher 2024).

Over the last several decades, these patterns have proven remarkably enduring and increasingly pronounced. The leftward shift has continued across changes in political leadership, economic fluctuations, and cultural currents. Unlike many other professions, which experienced periodic ideological rebalancing, the academy has drifted steadily to the Left (Kagan, Frake, and Hurst 2025). Regardless of the latest presidential election, shifts into and out of war, rising levels of right-wing populism, or changes of public opinion, the ideological composition of professors over the last several decades has become more and more liberal.

There are a variety of reasons one might lament all this ideological compression and imbalance: for the lack of professional networks afforded to conservative students; for the sense of alienation that can take hold of those who do not conform to liberal orthodoxies; and for the separation generated between universities and the larger societies in which they operate. For our purposes, though, the problem with ideological imbalance is the forfeiture of an essential, structural antidote to individual biases that degrade the research enterprise, undermine the search for truth and knowledge, and violate the public trust.

It isn't difficult to see how this happens. Ideological imbalances narrow the intellectual range of the university, not through overt censorship, but through shared assumptions that go unchallenged, dissenting perspectives that quietly disappear, and curricula shaped by prevailing orthodoxies. Over time, this ideological sorting becomes self-reinforcing—deterring contrarian thinkers from entering the profession and subtly pressuring those already inside it to mute disagreement, hedge conclusions, and signal conformity.

But most importantly, for present purposes, ideological homogeneity increases the odds that biases go unchecked—even unnoticed. When faculty share the same priors, the safeguards of critical inquiry erode. What emerges is not simply research colored by latent political views or moral values, but an intellectual culture in which political commitments are mistaken for objectivity itself. For when individual biases positively correlate with one another, as they do in intellectually homogenous communities, systemic distortions take hold and claims of indoctrination begin to gain real purchase.

With the compromise of the research enterprise comes a corresponding erosion of public trust in universities themselves. A deeper concern takes hold: that research universities, in becoming ideologically uniform, have withdrawn into enclaves of consensus—prizing belonging over argument and a specific sense of virtue over truth-seeking. Efforts to hire additional conservative faculty, in those rare instances when undertaken, routinely founder (Shields 2018).

This was never the compact that research universities forged with the public, one that conferred magnificent job protections, professional autonomy, and financial backing. With trust now falling, universities rush to reassure: that important research continues, that vital discoveries are being made, that—as my home institution quite rightly affirms—research saves lives. However, having failed to attend to the internal cultures and practices that buttressed their long ideological drift, many universities are playing defense. Unless they confront their internal cultures and practices directly, they will remain intellectually diminished, politically isolated, and stewards of an increasingly precarious research enterprise.

Conclusion

In the last several years, nearly every aspect of the academic enterprise has come under scrutiny or attack: hiring and admissions policies, academic freedom, accreditation practices, the nonprofit status of institutions, indirect cost recovery, and universities' eligibility for government grants and contracts. Under the second Trump administration, in particular, many of these pillars have been targeted for rollbacks and outright cuts that amount to billions of dollars in potential losses for research universities (Green 2025; Lemann 2026; Chronicle Staff 2025; Bender, Blinder, and Schmidt 2026). The larger project to delegitimize higher education continues apace.

There is plenty of blame to go around. Much of it lies with the political missionaries leading these attacks. They trade in anti-intellectualism, wield "accountability" as a cudgel against academic freedom, and seek not to reform higher education, but to discredit and dismantle it. Steven Pinker (2025), who is no great apologist for research universities, recently lamented the Trump administration's utter lack of "proportionality in dealing with [the nation's] educational and cultural institutions." He concluded his widely cited essay with this warning: "To cripple the institutions that acquire and transmit knowledge is a tragic blunder and a crime

against future generations." Universities are right to resist this onslaught—for the principles that they uphold, for the vital work they conduct, and for democracy itself.

Research universities, however, are not blameless. They have compromised the public's trust in at least two ways: by putting academic inquiry to the service of political advocacy and by cultivating a professoriate whose political commitments tilt overwhelmingly in one direction. These are not recent or accidental lapses. Despite longstanding admonition, and despite numerous opportunities to course correct, universities committed these breaches repeatedly, blithely, and under the naïve assumption that public support would be unwavering, that trust was assured. It was not.

The effect of either one of these violations isn't to break the public trust, all at once. A snapping sound doesn't come from any single break. Rather, each has the effect of further taxing, straining, and depleting available reserves. Losses occur incrementally, but there may exist critical thresholds that, once passed, make it nearly impossible to recover the public's prior willingness to support research universities.

Of course, not all institutions, divisions, or departments are equally culpable. Clark Kerr's famous observation that universities are best understood as "multiversities" is even more apt today than when he originally made it in 1963. Generalizations about them quickly run afoul of reality. Plenty of researchers adhere to the public compact and rightly deserve its trust. By trafficking in caricatures, many of the loudest critics of research universities misrepresent the intellectual seriousness, methodological rigor, and diversity of thought—if not ideological, then methodological and disciplinary—that continue to define much of the academic enterprise.

The misrepresentations of others, however, do not absolve those of us who work in higher education of responsibility. It would be a mistake to meet present critiques exclusively with defiance, deflection, and denial. If the research enterprise is to reclaim its standing, a reckoning must begin from within universities. We must gather ourselves, lower our guard, acknowledge our failings, return to first principles, and demonstrate—through conduct, not just declaration—that the disinterested pursuit of truth, intellectual pluralism, and honest transmission of knowledge remain central to our mission. Either we lead such reform ourselves, or we surrender the task—and the future of our research institutions—to those who would gladly see them undone.

■ *Oliver Buntin provided valuable research assistance. For feedback on earlier versions of the essay, the author is grateful to Peter Arcidiacono, Chris Celenza, Anthony Fowler, Ken Goldstein, Rick Hess, Jason Merchant, Bhaven Sampat, Sadia Sindhu, and the editors at the Journal of Economic Perspectives.*

References

Abbot, Dorian, A. Bikfalvi, April L. Bleske-Rechek, W. Bodmer, P. Boghossian, C.M. Carvalho, Joseph Ciccolini, et al. 2023. "In Defense of Merit in Science." *Journal of Controversial Ideas* 3 (1): 1–25.

Abrams, Samuel, and Karen Grigorian. 2018. "Administrators' Views on Campus Life, Diversity, and Politics." NORC at the University of Chicago. https://www.norc.org/research/projects/administrators-views-on-campus-life-diversity-and-politics.html.

Azoulay, Pierre, Joshua S. Graff Zivin, Danielle Li, and Bhaven N. Sampat. 2019. "Public R&D Investments and Private-Sector Patenting: Evidence from NIH Funding Rules." *Review of Economic Studies* 86 (1): 117–52.

Balafoutas, Loukas, Jeremy Celse, Alexandros Karakostas, and Nicholas Umashev. 2025. "Incentives and the Replication Crisis in Social Sciences: A Critical Review of Open Science Practices. "*Journal of Behavioral and Experimental Economics* 114: 102327.

Bender, Michael C., Alan Blinder, and Michael S. Schmidt. 2026. "Investigations and a Billion-Dollar 'Shakedown': How Trump Targeted Higher Education." *New York Times*, January 21. https://www.nytimes.com/2026/0½1/us/politics/trump-higher-education-2025.html.

Borjas, George J., and Nate Breznau. 2024. "Ideological Bias in Estimates of the Impact of Immigration." NBER Working Paper 33274.

Brady, Henry E., and Thomas B. Kent. 2022. "Fifty Years of Declining Confidence and Increasing Polarization in Trust in American Institutions." *Daedalus* 151 (4): 43–66.

Chronicle Staff. 2025. "Tracking Trump's Higher-Ed Agenda." *Chronicle of Higher Education*, April 18. https://www.chronicle.com/article/tracking-trumps-higher-ed-agenda.

Clune, Michael W. 2024. "We Asked for It: The Politicization of Research, Hiring, and Teaching Made Professors Sitting Ducks." *Chronicle of Higher Education*, November 18. https://www.chronicle.com/article/we-asked-for-it.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2002. "Links and Impacts: The Influence of Public Research on Industrial R&D." *Management Science* 48 (1): 1–23.

Council for Advancement and Support of Education. 2026. *CASE Insights on Voluntary Support of Education (United States): 2025 Key Findings*. Council for Advancement and Support of Education.

Cole, Jonathan R. 2009. *The Great American University: Its Rise to Preeminence, Its Indispensable National Role, Why It Must Be Protected*. PublicAffairs.

Daniels, Ronald J. 2021. *What Universities Owe Democracy*. Johns Hopkins University Press.

Druckman, James, Alice Brocheux, Pauline Gordula, Hope E. Marsh, Dot Sawler, Daniel Sun, and Yi-Fan William Zhu. 2026. "Public Opinion Trends in American Society: LEssons from Social Science Infrastructure." *PNAS Nexus* 5 (2).

General Social Survey. 2025. *General Social Survey Report Archives: Confidence in Education*. NORC at the University of Chicago. https://gssdataexplorer.norc.org/variables/CONEDUC (accessed August 6, 2025).

Green, Emma. 2023. "The Right Side of History: How Should Historians Respond to the Urgency of this Current Political Moment?" *New Yorker*, March 7. https://www.newyorker.com/news/annals-of-education/the-right-side-of-history.

Green, Emma. 2025. "Inside the Trump Administration's Assault on Higher Education." *New Yorker*, October 13, 2025. https://www.newyorker.com/magazine/2025/10/20/inside-the-trump-administrations-assault-on-higher-education.

Geiger, Roger L. 2004. *Research and Relevant Knowledge: American Research Universities since World War II*. Routledge.

Gross, Neil, and Solon Simmons. 2007. "The Social and Political Views of American Professors." Unpublished.

Gupta, Akhil, and Jessie Stoolman. 2022. "Decolonizing US Anthropology." *American Anthropologist* 124 (4): 778–99.

Hertzog, Matthew. 2017. *Protections of Tenure and Academic Freedom in the United States: Evolution and Interpretation*. Palgrave Macmillan

Hess, Frederick M., and Riley Fletcher. 2024. "Top Public Policy Programs Have Almost No Conservative Faculty." Manhattan Institute, November 21. https://www.aei.org/wp-content/uploads/2024/11/top-public-policy-programs-have-almost-no-conservative-faculty.pdfpage8.pdf.

Higher Education Research Institute. 2016. *The HERI Faculty Survey: 2016 Results*. University of California, Los Angeles.

Honeycutt, Nathan. 2024. "Confidence in Colleges and Universities Hits New Lows, per FIRE Polls." Foundation for Individual Rights and Expression, June 11. https://www.thefire.org/news/confidence-colleges-and-universities-hits-new-lows-fire-polls.

Honeycutt, Nathan. 2025. "The Politics of University Faculty." Unpublished.

Humphrey, Justin. 2025. "The Pell Grant Program: Funding and Current Shortfall." Presentation at the American Association of State Colleges and Universities, April 29. https://www.cbo.gov/system/files/2025-04/61366-Pell-Grant-Program.pdf.

Jelveh, Zubin, Bruce Kogut, and Suresh Naidu. 2024. "Political Language in Economics." *Economic Journal* 134 (662): 2439–69.

Jones, Jeffrey M. 2026. "Confidence in U.S. Higher Education Slips Back Slightly." *Gallup*, July 13. https://news.gallup.com/poll/712322/confidence-higher-education-slips-back-slightly.aspx.

Kagan, Max, Justin Frake, and Reuben Hurst. 2025. "VRscores: A New Measure and Dataset of Workforce Politics Using Voter Registrations." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.5104795.

Kerr, Clark. 1963. *The Uses of the University*. Harvard University Press.

Khaitan, Tarunabh. 2022. "On Scholactivism in Constitutional Studies: Skeptical Thoughts." *International Journal of Constitutional Law* 20 (2): 547–56.

Langbert, Mitchell. 2018. "Homogenous: The Political Affiliations of Elite Liberal Arts College Faculty." *Academic Questions* 31 (2): 186–97.

Langbert, Mitchell, and Sean Stevens. 2020. "Partisan Registration and Contributions of Faculty in Flagship Colleges." *National Association of Scholars*.

Lee, Stephanie. 2026. "The War for Sociology: Should scholars be activists? A famously left-leaning field is divided." *Chronicle of Higher Education*.

Lemann, Nicholas. 2026. "The Unmaking of the American University." *New Yorker*, March 9. https://www.newyorker.com/magazine/2026/03/16/the-unmaking-of-the-american-university.

Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.

Macedo, Stephen, and Frances Lee. 2025. *In Covid's Wake: How Our Politics Failed Us*. Princeton University Press.

Manzi, James. 2026. "The Ideological Orientation of Academic Social Science Research 1960–2024." *Theory and Society* 55 (2): 25.

Merton, Robert K., ed. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.

Mowery, David, Richard Nelson, Bhaven Sampat, and Arvids Ziedonis. 2001. "The Growth of Patenting and Licensing by U.S. Universities: An Assessment of the Effects of the Bayh-Dole Act of 1980." *Research Policy* 30 (1): 99–119.

Nagel, Thomas. 1986. *The View from Nowhere*. Oxford University Press.

National Science Board. 2025. *Science and Engineering Indicators*. NSB-2025-7. https://ncses.nsf.gov/pubs/nsb20257/academic-r-d (accessed June 22, 2026).

Nature Human Behaviour. 2022. "Science Must Respect the Dignity and Rights of All Humans." *Nature Human Behaviour* 6: 1029–31.

Pinker, Steven. 2025. "Harvard Derangement Syndrome." *New York Times*, May 23. https://www.nytimes.com/2025/05/23/opinion/harvard-university-trump-administration.html.

Primo, David M. 2026. *Faculty Ideology: Measuring Faculty Viewpoint Diversity Using Campaign Contribution Data*. Foundation for Individual Rights and Expression.

Reichman, Henry. 2021. *Understanding Academic Freedom*. Johns Hopkins University Press.

Rosenberg, Nathan, and Richard R. Nelson. 1994. "American Universities and Technical Advance in Industry." *Research Policy* 23 (3): 323–48.

Rubin, Ashley. 2025. "Normativity Is Not a Replacement for Theory." *Theory and Society* 1–56

Savolainen, Jukka. 2025. "Left-Wing Bias Is Corrupting Sociology." *Chronicle of Higher Education*, October 22. https://www.chronicle.com/article/left-wing-bias-is-corrupting-sociology.

Scott, Joan Wallach. 2019. *Knowledge, Power, and Academic Freedom*. Columbia University Press.

Scheufele, Dietram A., Nicole M. Krause, Isabelle Freiling, and Dominique Brossard. 2020. "How Not to Lose the COVID-19 Communication War." *Slate*, April 17. https://slate.com/technology/202%4/covid19-misinformation-science-communication.html.

Shields, Jon. 2018. "The Disappearing Conservative Professor." *National Affairs* 37: 138–50.

Shields, Jon A., and Joshua M Dunn Sr. 2016. *Passing on the Right: Conservative Professors in the Progressive University*. Oxford University Press.

Slaughter , Sheila, and Larry Leslie. 1997. *Academic Capitalism: Politics, Policies, and the Entrepreneurial University*. Johns Hopkins University Press.

Veblen, Thorstein. 2015 (1918). *The Higher Learning in America: A Memorandum on the Conduct of Universities by Business Men*. Johns Hopkins University Press.

Wright, Graham, Shahar Hecht, and Leonard Saxe. 2025. "Ideology in the Classroom: How Faculty at US Universities Navitgate Politics and Pressure Amid Federal Pressure Over Viewpoint Diversity and Anti-Semitism." *Brandeis University*.

Journal of Economic Perspectives—Volume 40, Number 4—Fall 2026—Pages 73–94

The Role of Private Debt in the Financial Ecosystem

Victoria Ivashina

All firms require financing, which at a high level can take the form of either debt or equity. Historically, equity has been categorized as either publicly traded or privately held, with the latter typically owned by specialized private equity firms. A similar distinction has existed in debt markets: academic economists traditionally divided debt into public debt (bonds) and private debt, referring primarily to bank lending. However, as commercial debt markets have evolved, "private debt" has come to denote debt originated by, and typically held on the balance sheets of, *alternative* financing vehicles—namely, nontraditional structures such as private credit funds and business development companies. Thus, nonpublic debt now consists of two broad segments: bank-originated lending and private debt.

In the 15 years following the 2007–2009 global financial crisis, the private debt market evolved from a niche investment strategy into a major segment of global debt markets. This development raises important questions about the implications for traditional debt markets and the banking sector in particular, as well as broader financial stability. The goal of this paper is to provide an introduction to private debt, examine the drivers of its expansion, and frame the discussion around potential risk build-up in the economy.

In what follows, I begin by providing background on the evolution of private debt and outlining the key factors that drove the emergence of the modern private debt market after the global financial crisis of 2008–2009. This is followed by a deep

■ *Victoria Ivashina is Professor of Business Administration, Harvard Business School, Boston, Massachusetts, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Her email address is vivashina@hbs.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251482.

dive into the unique features of private debt and the reasons private creditors are able to provide these compared with other sources of capital. I then introduce the two funding structures common in the private debt space: (1) private equity–style credit funds that raise illiquid commitments from large institutional investors; and (2) evergreen structures oriented toward retail investors, which in the United States are typically called "business development companies."

As highlighted by the surge in subprime lending prior to the global financial crisis of 2007–2009 and, more recently, the collapse of Silicon Valley Bank in 2023, it is prudent to monitor the rapid growth of isolated segments of the financial market—especially when accompanied by high levels of debt and structural complexity. Unsurprisingly, the expansion of private debt has drawn increasing scrutiny (for example, Cortes et al. 2024; Avalos, Doerr, and Pinter 2025; Financial Policy Committee 2025). Thus, I conclude with a discussion of the broader implications of the rise of private debt for financial stability. To be upfront, while my perspective is nuanced, I view private debt overall as a positive development—and one that is here to stay.

Evolution of the Private Debt Market

To date, much of the growth in private debt has been concentrated in the high-yield—or highly leveraged—segment of the corporate debt market.[1] The bulk of private debt consists of loans with high credit risk typically secured by some form of collateral. A representative use of private debt is the financing of a leveraged buyout (LBO), in which a private equity sponsor acquires a company using a combination of debt and equity.

More broadly, the evolution of high-yield debt markets, including the rise of private debt, has been closely tied to the history of private equity, particularly leveraged buyouts. Over the past 50 years, the buyout industry has gone through three major growth waves, each expanding the footprint of private equity in the economy and each closely linked to the development of a distinct segment of the risky debt market: first high-yield bonds, then leveraged loans, and most recently private debt.[2] While leveraged loans and, especially, high-yield bonds primarily serve larger companies, the growth of private debt has extended high-yield corporate credit to middle market companies. The usual dividing line is that core middle market refers to firms that have EBITDA (earnings before interest, taxes, depreciation, and amortization) of $25–75 million, while firms with larger EBITDA are classified as large-cap.

[1] As a reference point, over the past decade, loans backing buyout transactions have delivered gross-of-fees returns of roughly 9 percent per year (Sundar 2025).

[2] High-yield bonds are bonds rated "BB"/"Ba" or below. Leveraged loans include loans rated "BB"/"Ba" or below, as well as unrated loans that carry a comparable level of risk and, accordingly, offer similar interest rates.

As of the end of 2025, multiple sources estimate that the global private debt market has reached roughly $2.0 trillion in committed capital, up from negligible levels in 2008, with much of it concentrated in the United States. About one-third of this capital is committed by investors, but remains undeployed. Accordingly, in terms of outstanding amounts, the private debt market is now broadly comparable in size to the US leveraged loan market, at about $1.6 trillion, and the US high-yield bond market, at about $1.5 trillion.

Debt is an intrinsic feature of private equity financing. In a typical modern buyout transaction, specialized investment firms acquire a controlling position in a company with the goal of pursuing growth through operational changes, generally over an ownership horizon of about five years. Because such transactions involve acquiring a controlling stake, the required investment is usually substantial and is financed with a combination of equity and debt. Moreover, private equity funds typically target a fixed, leveraged gross return of around 20 percent. This return expectation is embedded in market practice and acts as an additional constraint on the types of financing used in buyout transactions. For a comprehensive introduction to leveraged buyouts and private equity, a useful starting point in this journal is Kaplan and Strömberg (2009).

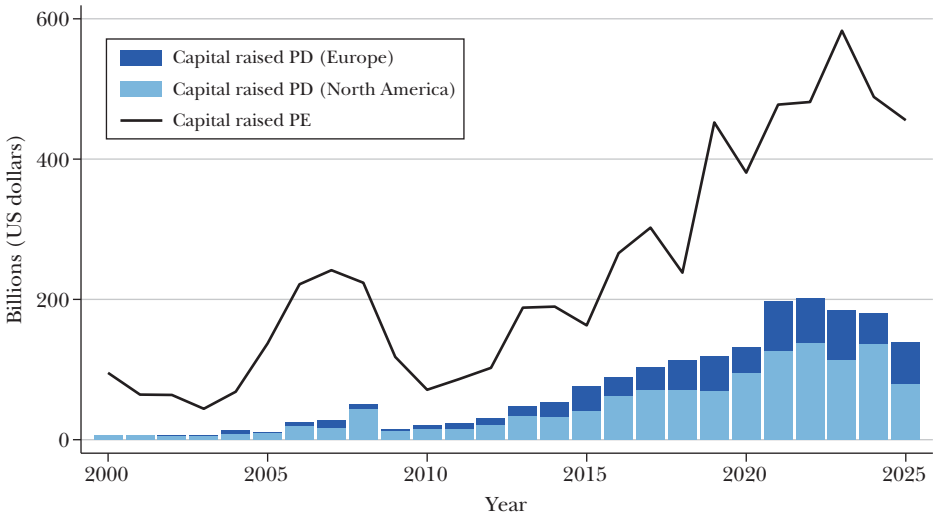
Figure 1 illustrates the growth of private debt fundraising since 2000 and compares it with the growth of private equity. It provides useful background for distinguishing between the two periods of the private debt market (before and after the 2007–2008 financial crisis) while also highlighting its connection to private equity market.

Private Debt 101

As the private equity industry experienced rapid growth in the early 2000s, increasing competition pushed firms to target smaller companies. However, the high-yield debt market serving these smaller, mid-sized firms remained relatively shallow and underdeveloped. In contrast, the bulk of the debt financing for larger buyouts was raised in the leveraged loan market—the non-investment-grade loan market funded by a broad base of institutional investors. For example, according to PitchBook Leveraged Data & Commentary database ("PitchBook" from now on), the average leveraged buyout loan in the US leveraged loan market in 2004 was approximately $354 million, representing roughly 44 percent of total buyout financing. In contrast, loans backing mid-sized buyouts were often too small to be syndicated for similar reasons that explain why smaller firms typically do not issue bonds: the process involves significant fixed costs related to rating, underwriting, and placement (for a description of the process, see Ivashina and Mylavarapu 2025).

At the same time, on-balance-sheet bank debt was an insufficient source of debt financing for these buyouts. Due to capital and supervisory requirements, banks typically provided only a small, most senior position in the capital structure when lending to buyouts. Figure 2 shows an illustrative capital structure of a middle market buyout before the 2007–2009 financial crisis. Moreover, in the late 1990s and early 2000s, driven by regulatory changes, the US banking sector underwent significant

Figure 1
Evolution of Private Debt and Private Equity Fundraising



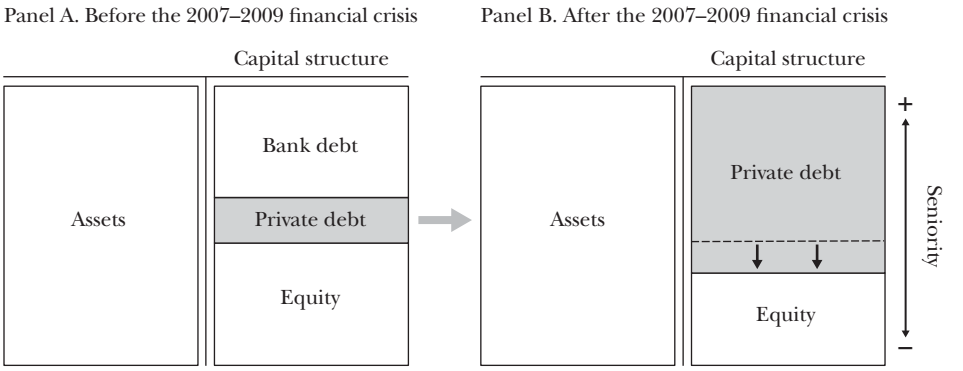
Source: Ivashina (2026).
Note: The figure is based on Preqin data. Private equity includes buyout and growth equity (PE) strategies, with the line representing fundraising in North America and Europe. Private debt (PD) fundraising excludes fund-of-funds strategies to avoid double counting, as well as distressed debt, which is a hostile strategy distinct from the rest of PD.

consolidation and expansion in scope. By some accounts, this marked the beginning of large banks' retreat from mid-market and small commercial lending, as they shifted their focus toward larger clients. These clients presented opportunities for banks to cross-sell fee-based services such as underwriting, merger and acquisition advisory, and loan syndication.

It was in this context—amid growing demand for mid-sized leverage and as banking activity increasingly focused on larger firms—that the first generation of private debt funds emerged. These were the so-called "captive" private debt funds. These funds became ubiquitous among mid-cap private equity managers in the years leading up to the 2007–2009 financial crisis. During this period, many large growth-equity and mid-market buyout firms began raising not only equity funds, but also debt funds designed to invest in the same deals. The common fundraising pitch was: "You already invest in equity with us—why not also invest in the senior, safer claims of the same companies?" The issue, of course, is that when things go wrong, there is an inherent conflict of interest between debt and equity holders that is very difficult to mitigate. That did not stop captive funds from making meaningful strides in fundraising.

Between the limited presence of private equity in the middle market before the 2007–2009 financial crisis and the conflicts of interest associated with captive credit funds, this early form of private debt remained modest in scale. After the

Figure 2
Representative Middle Market Buyout Structure



Source: Author's creation.
Note: This stylized figure depicts the role and representative forms of private debt financing used in buyouts during the early and later stages of private debt market development.

financial crisis, as a new generation of arm's-length private credit providers began to emerge alongside the rise in mid-cap leveraged buyout activity, the economics of captive debt funds became unattractive and they largely disappeared. Figure 2 illustrates the role of banks and the shift in the private debt market in financing middle market buyouts before and after the 2007–2009 financial crisis.

Emergence of Modern Private Debt Industry

Much of today's private debt industry traces its roots back to the 2007–2009 financial crisis. As pointed out above, the groundwork had already been laid in the years leading up to the crisis. Although private debt remained relatively small at the time, it was no longer unfamiliar to private equity sponsors or to institutional investors, who would become the core investor base for credit funds. While multiple factors have contributed to the growth of private debt since 2008, two primary forces stand out.

First, the banking sector emerged from the 2007–2009 financial crisis with tighter regulation and increased scrutiny. Among other things, the revised Leveraged Lending Guidance issued in 2013 articulated stricter supervisory expectations for bank-originated leveraged loans. Davydiuk, Marchuk, and Rosen (2024) provide a thorough account of the regulatory and structural changes that drove the retreat of traditional lenders from riskier segments, perhaps disproportionately in the middle market, which—unlike the leveraged loan market—has never been a significant source of fee generation for banks.

However, while post–financial crisis pressures on banks likely contributed to the rise of private debt, banks had not been significant providers of risky lending to the middle market to begin with (as shown in Figure 2). As a result, a simple credit provider substitution (that is, a credit supply explanation) cannot fully account for

the rapid expansion of risky, high-yield lending to middle market firms by private debt. Instead, understanding what was happening on the private equity side (that is, the demand for private credit) is essential.

The 2007–2009 financial crisis marked the beginning of an extended period of historically low interest rates across advanced economies. During the crisis, central banks around the world cut policy rates to near zero in an effort to stabilize financial markets and stimulate the economy, a move reinforced by unconventional monetary policies. In this environment, institutional investors with long-term liabilities significantly increased their target allocations to private asset classes, including private debt. This rise in target allocations fueled many years of a fundraising boom for these asset classes.

Ivashina and Lerner (2018) examine pension funds around the world and show that, across countries and fund sizes, there has been a clear shift toward private markets. In developed economies, allocations to private assets rose from 7 to 12 percent of assets under management between 2008 and 2017. In emerging markets, allocations increased from less than 1 percent to nearly 7 percent over the same period. Much of this capital flowed into US private equity and debt funds. It was a structural shift in allocation policies and a global phenomenon that cannot be explained by country-level factors.

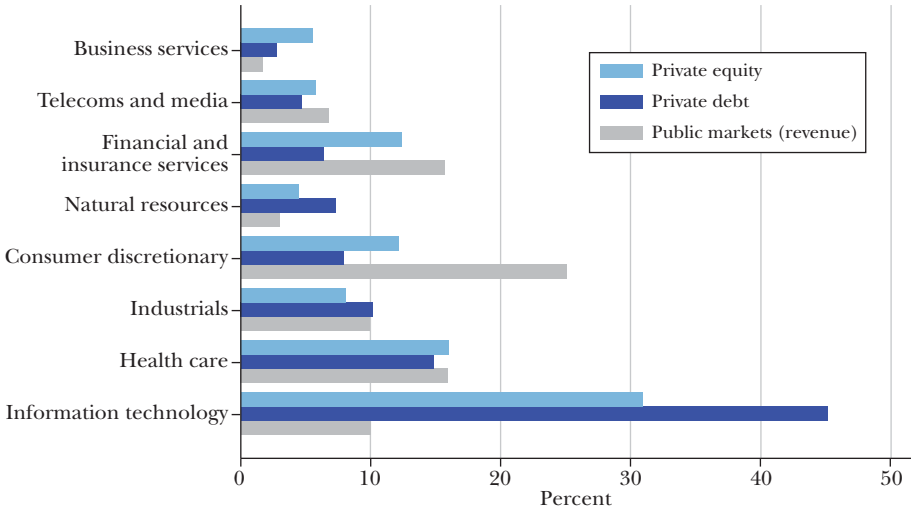
The economic rationale behind the massive inflow of capital into private space is nuanced and is beyond the scope of this paper. But between 2013 and 2023:II, private assets grew at 14 percent compound annual growth rate per year. Over the same period, aggregate private debt assets expanded 25 percent per year, albeit starting from a much smaller base (Bain & Company 2024).

In sum, after the 2007–2009 financial crisis, the low interest-rate environment has been associated with significant capital inflows into the alternatives space, facilitating fundraising for both private equity and private debt. At the same time, banks have been more reluctant than ever to use their balance sheets to lend in the high-yield space. This combination has led to the emergence of large, arm's-length private debt providers to support private equity activity.

As of the end of 2025, the private debt market remains heavily concentrated in "direct lending," provided either through private equity-style credit funds or retail-oriented structures such as business development companies, and it continues to serve primarily the private equity industry. PitchBook—a widely used database for tracking investments of business development companies—does flag buyout transactions, but this flag tends to refer only to the initial buyout deal. Any interim refinancing or follow-on lending within that five-year window may not be marked as buyout-related, even though the borrower remains under private equity ownership. As a result, the data may understate the full extent of private equity involvement in the private debt market.

To address this, I used standard assumptions about the typical life cycle of a private equity deal to be able to attribute loans to companies backed by private equity. My estimates suggest that approximately 60 percent of business development companies' investments—whether measured by volume or by number of loans—support

Figure 3
Industry Distribution of US Private Debt and Private Equity Deals, 2020–2024



Source: Ivashina (2026).
Note: Compiled using Preqin and Compustat data. Calculated on a dollar volume basis. Energy & Utilities, Real Estate, and Unknown sectors are omitted from the figure, but included in the calculations of total volumes.

private equity activity. This figure is likely to be conservative. A survey by Block, Jang, Kaplan, and Schulze (2024), conducted in 2021, indicates that on average 78 percent of US private credit funds in their sample were backing private equity activity (with a median of 85 percent). Conversations with investment committee members at several of the largest business development companies suggest that the actual share is closer to 85 percent.

The connection between private equity and private debt is also evident in the industry distribution of deals. Figure 3 shows that 45 percent of credit fund transactions by dollar volume were concentrated in information technology, with the bulk directed toward cloud computing, followed by mobile applications and artificial intelligence–related businesses. Information technology was likewise the dominant industry focus for private equity, corresponding to roughly one third of deal volume. In contrast, over the same period, the sample of US publicly traded companies was dominated by the consumer discretionary sector on a revenue basis and by the financial and insurance sector on an asset basis (the latter is reported in the Supplemental Appendix available on the JEP website).

Unique Features of Private Debt

As with other high-yield corporate credit markets, private debt primarily involves lending against enterprise value (versus the value of a fixed asset); this type

of lending is referred to as "cash flow credit." In practical terms, this means lending against forecasted earnings before interest, taxes, depreciation, and amortization or another measure of enterprise value. To manage credit risk—that is, the risk of not recovering promised payments in full or on time—all cash-flow lenders rely on a system of credit screening and ongoing monitoring.

Leveraged loans and high-yield bonds, however, depend on being widely distributed and are actively traded.[3] Given its smaller size, private credit is funded from the balance sheet of the private credit provider and does not have an active secondary market. That is, private debt is primarily originated to hold. This, in turn, underpins how private lenders manage credit risk and some other distinctive features of this type of credit.

Different Approach to Information Collection

Similar to private equity, private debt involves proactive sourcing (versus a borrower reaching out to the lender). For example, in private equity, one often hears that the average time from first contact to deal closure is about four years, during which the investment firm actively follows the target company and engages with its management. This relationship management is an intrinsic part of deal origination and due diligence in private equity. A very similar multiyear dynamic takes place in the modern private debt space although the relationship focus is on the private equity managers (who manage funds from which buyout investments are made), and due diligence is conducted not only on the underlying deal, but also on the private equity sponsor, their fund, and the broader private equity industry.[4]

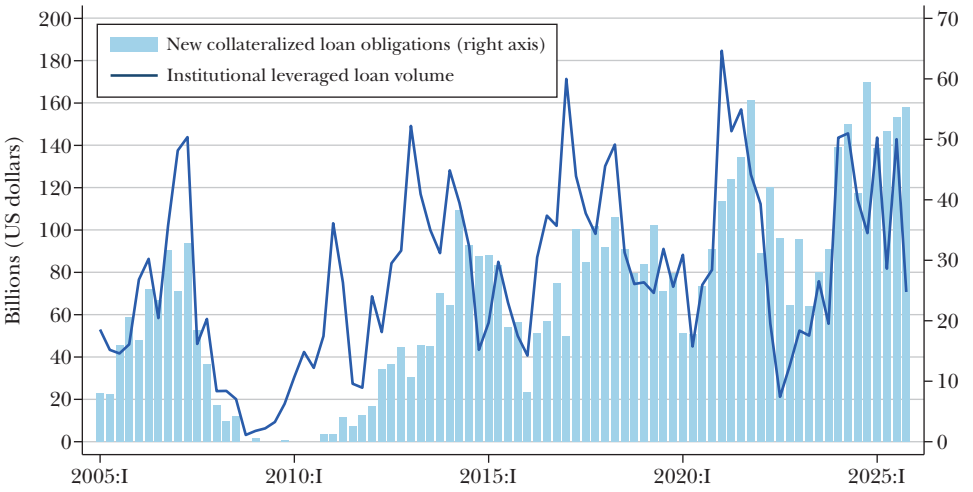
This costly screening and monitoring function does not appear necessary in the context of bank middle-market lending to buyouts, where banks typically hold only a small, very safe exposure. Leveraged loans, on the other hand, are risky, but they are originated by banks and funded primarily by a broad syndicate of nonbank institutional investors, with collateralized loan obligations (CLOs), a type of securitization vehicle, serving as the dominant source of capital.[5] (Within the high-yield space, leveraged loans are often referred to as "bank debt," though the term more accurately reflects "bank-originated debt.") Because leveraged loans are funded primarily by nonbanks, they are highly dependent on market conditions, much like bond underwriting.

[3] A common misconception is that any nonbank financial intermediary (NBFI) holding corporate loans qualifies as private debt. In practice, large-cap collateralized loan obligations and leveraged loans more broadly are not considered part of the private debt market.

[4] Private debt lenders also invest in developing in-house industry expertise, though it is unlikely to be superior to that of large banks and therefore is not a substantial differentiator.

[5] Securitization is the process of funding a *diversified* pool of *risky* claims through the issuance of *safe debt* and equity. Traditionally, such debt has been raised by issuing public bonds (for example, in mortgage securitization), though it can take other forms. Collateralized loan obligations are a specific example of securitization, in which a pool of leveraged loans is financed by issuing safe debt and equity.

Figure 4
Leveraged Loan Market: Collateralized Loan Obligations Issuance and Total Institutional Volume



Source: Ivashina (2026).
Note: The figure shows new collateralized loan obligations issuance and contemporaneous total institutional leveraged loan volume, which corresponds to funding of newly issued loans. The figure is compiled using PitchBook LCD data.

Figure 4 illustrates this point. The bars represent new issuance of collateralized loan obligations. The line shows the volume of newly issued leveraged loans funded by institutional money. Clearly, issuance of new collateralized loan obligations is highly sensitive to market conditions—it came to a complete halt in 2007–2009, dropped sharply during the COVID-19 pandemic, and declined again amid rising inflation and heightened recession risk that followed. Figure 4 also shows significant fluctuations in loan issuance volume that closely mirror movements in issuance of collateralized loan obligations. This stands in contrast to private creditors, whose on-balance-sheet funding and on-demand access to committed but uncalled capital (for credit funds) allows for a fundamentally different approach to operations.

In sum, bank bilateral originate-to-hold lending in the middle market is not as risky and therefore does not require the costly proprietary information processing developed by private creditors. At the same time, bank originate-to-distribute leveraged loans are risky but allow for risk diversification through distribution and, in addition, are unlikely to justify investment in actively managed, long-term relationships when funding conditions are ultimately externally determined.

Beyond operationalizing high-yield lending to middle market firms, private debt has also ventured into areas that are largely new to debt, including recurring revenue lending (common in the technology industry) and client acquisition financing (common among start-ups)—segments that were historically funded

primarily by equity. This is reflected in Figure 3, where many of the technology companies receiving private debt financing are fast-growing but not yet profitable software-as-a-service (SaaS) businesses.

While private debt has certainly developed expertise in this area, its dominance likely stems less from superior information processing and more from the fact that (unlike banks) it is not constrained to lending only to earnings before interest, taxes, depreciation, and amortization–positive businesses (Chernenko, Erel, and Prilmeier 2022). This flexibility allows private debt lenders to underwrite and contract on other measures of enterprise value, such as recurring revenue. By contrast, the bank-originated leveraged loan market remains EBITDA-based, with a typical loan size in the range of 4.5 to 5 times EBITDA—a structure that is not a feasible funding option for many highly valuable but not yet profitable software companies.

Private Debt Loans Versus Bank-Originated Loans

The way private creditors collect information and originate credit helps explain how private debt is doing something that has not previously been done on a significant scale: providing arm's-length high-yield debt to middle market firms. This feature is what makes private debt a unique segment of the broader high-yield corporate debt space. On the margin, however, there is some overlap between borrowers who can tap either the private debt market or the leveraged loan market.[6] Thus, larger or fast-growing mid-cap firms with stable and positive earnings before interest, taxes, depreciation, and amortization often consider two alternative debt financing structures: one backed by a private debt provider and the other funded through the leveraged loan market.[7] In this subsection, I outline the core features that tend to be weighed when such a direct comparison is applicable.

First, private debt is illiquid, on-balance-sheet lending in which the borrower typically interacts with a *single and stable counterparty*. Paired with the deep pools of "dry powder" capital available to credit funds, this clarity can make private debt a preferred source of financing. At closing, it offers borrowers speed, certainty of terms, and confidentiality, all of which can be valuable in the highly competitive private equity context. The same factors and dynamics apply when a loan needs to be refinanced or restructured (Bolton and Scharfstein 1996).

In contrast, leveraged loans and public bonds are widely distributed among investors and actively traded in the secondary market. Placing either instrument requires obtaining credit ratings and conducting a roadshow to market the deal to potential investors—a process that can take up to two months, even in favorable

[6]The median borrower in the two markets continues to be very different: according to PitchBook, between 2020 and 2024, the median buyout loan funded by private creditors was about $260 million, compared with $1.2 billion for the median buyout loan in the leveraged loan market.

[7]See Ivashina and Mylavarapu (2025) for a representative real-life illustration of buyout financing decisions with multiple debt package options, including representative term sheets.

market conditions. During this time, the borrower remains exposed to both deal risk and market risk, as conditions may shift adversely before the transaction is completed.

Note that on-balance-sheet lending by a single or limited group of creditors is typically associated with stricter contractual terms. However, this is an endogenous outcome, as renegotiation in this setting is easier.

Second, where other options are available, private debt has generally been somewhat *more expensive*. PitchBook indicates that, between 2018 and 2024, average interest rates on leveraged loans were consistently lower than those on private credit loans of comparable seniority and collateralization, with the difference in the range of 1.4 percent to 2.2 percent. However, this statistic likely overstates the gap, as it is conditional on the type of debt issued and therefore does not account for selection, nor for the fact that the leveraged loan market can be subject to significant fluctuations. It also does not adjust for leverage, and private creditors tend to tolerate slightly higher levels of debt than bank-originated leveraged loans. This gap has narrowed over time, likely due to rising competition and ongoing consolidation within the private debt market, which has reduced operating costs.

It is common to hear the argument that bank credit is cheaper than private debt because of banks' funding costs: bank deposits are protected by deposit insurance and, at least for the largest banks, there may be benefits from an implicit bailout guarantee. This argument is misleading, as banks' on-balance-sheet lending is not directly comparable to private debt. The relevant comparison is the leveraged loan market (bank-originated debt), which, as discussed above, is heavily dependent on market conditions, and this is what dictates the cost of the leveraged loans.

The differences in funding costs between direct lending and "bank lending" in the large-cap high-yield space stem primarily from diversification and liquidity, rather than banks' cost of funding. In the case of a direct lender, a $200 million loan would enter the balance sheet of private creditors. In the case of a leveraged loan, the loan is distributed among a range of institutional investors, with each individual creditor holding only about a $2–$5 million position. Moreover, liquidity in the leveraged loan market is comparable to that of the bond market. In contrast, secondary sales by private creditors are rare.

Third, structurally, while private debt—like leveraged loans—tends to be senior secured, *it can also take the form of equity* or, more commonly, a custom hybrid of debt and equity. According to PitchBook data, as of the end of 2024, 86 percent of aggregate business development company holdings (by book value) consisted of senior secured debt, with an additional 3 percent in second-lien debt. Subordinated debt accounted for about 3 percent, while equity represented around 8 percent. Importantly, in situations that call for flexibility, such as borrowers emerging from or currently in financial distress, equity features become far more prominent, often through warrants or debt-to-equity conversion. This structural flexibility, combined with a single-counterparty relationship, provides restructuring optionality that is typically unavailable in the bank-originated market due to institutional constraints and the lack of coordination among a dispersed creditor base.

Relatedly, using PitchBook data, I estimate that in 2024, 17.5 percent of loans held by business development companies included payment-in-kind provisions on a value-weighted basis, and 11.4 percent on an equal-weighted basis.[8] Typically, a borrower is required to make quarterly cash interest payments. The payment-in-kind feature allows the borrower to defer these payments by adding the interest to the loan's principal.

Fourth, there is one important area where banks continue to play a pivotal role: the issuance of *revolving credit*. Broadly speaking, there are two main types of commercial loans: term loans and revolving lines of credit. Term loans are drawn at origination and are terminated once they are repaid, much like a mortgage or a student loan. Revolving lines of credit, by contrast, can be drawn up to an authorized amount and repaid multiple times until maturity, similar to a credit card. Most companies maintain a revolving line of credit, which they use to manage working capital or finance small acquisitions and investments.

While a revolving line remains unused, the borrower pays only a small, fixed fee to keep it open. From the creditor's perspective, this represents a challenge, as most individual lines remain undrawn for long periods, generating little compensation while remaining callable at any time. Making the economics of revolving lines viable therefore requires both scale and the diversification that comes with it, as well as sophisticated liquidity management. Banks—especially deposit-taking banks—have a clear advantage in issuing revolving lines, a central insight of Kashyap, Rajan, and Stein (2002).

Given that revolving credit lines are a necessity, how is this addressed in the private debt space? In some cases, borrowers obtain revolving lines directly from banks, in which case private debt is subordinated to the balance on the revolving line. More frequently, private lenders themselves obtain a revolving line from banks, which is then used to fund the revolving lines they issue on the asset side. However, revolvers continue to be a "loss leader" for most private creditors, and their volume remains modest.

Funding Structures of Private Debt Vehicles

To discuss the financial stability implications of the growth of private debt, it is essential to first understand its funding structure. The market for private debt vehicles is currently divided between two core funding structures. One consists of traditional closed-end funds that primarily serve institutional investors. The other includes structures designed for retail investors, with business development companies being the dominant form in the United States. As of 2025, roughly two-thirds of

[8]These estimates are likely a lower bound, as for at least some business development companies, the presence of payment-in-kind provisions can only be observed in Securities and Exchange Commission filings or in PitchBook data when they are actually used. I classify a loan as having a payment-in-kind feature if it was utilized at any point prior to the end of 2024.

total private debt capital was managed by credit funds, with the remaining one-third in business development companies.

Credit Funds

Credit funds are normally structured in a way similar to private equity funds; that is, they are closed-end limited partnerships. These funds generally have a finite lifespan, with managers expected to return capital within ten years of the fund's launch. Funding is provided primarily by institutional investors with long-term liabilities—pension funds, insurance companies, and, to a lesser extent, endowments, sovereign wealth funds, and family offices. Capital is committed upfront and locked in for a specified time horizon, but is drawn down by the fund manager only as investment opportunities arise. To maintain business continuity, new funds are raised on an overlapping basis.

Due to the locked-up nature of their capital commitments, private debt funds tend to maintain substantial funding capacity even during economic downturns.[9] For example, in 2024, private debt "dry powder"—that is, committed but not deployed *equity* capital—exceeded half a trillion dollars.

Credit funds can employ leverage, most commonly through bank credit lines secured against existing assets. A survey by Block, Jang, Kaplan, and Schulze (2024) indicates that almost all US funds in their sample use some leverage, with the average fund reporting a 1:2 debt-to-equity ratio (median of 1:3). The same study finds that 60 percent of European credit funds do not use leverage at all. From a financial stability perspective, the central point is that the bulk of private credit equity capital and, importantly, uncalled capital commitments is locked in over long horizons. As a result, like private equity funds, existing credit fund structures are relatively insensitive to market fluctuations.

Business Development Companies

Business development companies are legal structures under US law. They are a form of "retail vehicles," meaning that they allow a wide range of households to gain exposure to the private debt asset class. According to *Raymond James BDC Weekly Insight*, average institutional ownership of business development companies was only 22.2 percent as of the end of April 2026.

Unlike credit funds, business development companies do not have unused equity capital reserves and must provide some form of liquidity to their equity investors. Accordingly, business development companies are either publicly traded or offer periodic (typically quarterly) liquidity to investors. Private business development companies are generally "gated" by default, limiting redemptions to about 5 percent of net asset value, although these limits can be—and in some cases have been—lifted.

[9]Defaulting on capital commitments is highly punitive for investors in private equity and private debt funds. As a result, investors are far more likely to sell their commitments on the secondary market at a discount than to default outright.

Business development companies are subject to regulatory minimum capital requirements set by the Securities and Exchange Commission. Historically, this meant maintaining a 1:1 debt-to-equity ratio, regardless of the risk profile of their loan portfolios. In 2018, this requirement was relaxed to a 2:1 ratio, allowing greater leverage. That said, the median business development company still operates at about a 1:1 ratio—higher than that of credit funds but significantly lower than that of banks.

Chernenko, Ialenti, and Scharfstein (2025) provide a comprehensive analysis of business development companies using a bank regulatory framework, concluding that if these firms were regulated as banks, their capitalization would be equivalent to approximately 36 percent of risk-weighted assets. The same paper shows that about 20 percent of their funding is drawn from bank lines of credit, and 30 percent is funded through bonds.

I will expand on issues of financial stability risk in more detail in the next section, but the liability structure outlined above already points to three key elements to consider. First, the reliance of business development companies on market-based sources of capital may limit their ability to lend during market downturns, as access to funding tightens. Second, business development companies have nontrivial links to the banking sector, with roughly 20 percent of their funding coming from bank-provided revolving credit lines. These connections create potential channels of contagion between business development companies and traditional banks. However, these revolving credit lines also act as a form of liquidity hedge for business development companies, offering some protection against deteriorating market conditions (Acharya, Cetorelli, and Tuckman 2026). Third, business development companies are subject to redemption pressures, and even if liquidity gates are upheld, prolonged pressure could be an important constraint on new loan origination over the medium and long run.

Implications for Financial Stability

In what follows, I explore key themes and concerns that have emerged around the potential systemic risks associated with the growth of private debt.

Evaluating Potential Structural Vulnerabilities of Private Creditors

Much like banks, private lenders rely on proprietary screening and monitoring technology. As a result, it is time-consuming and costly for the borrower to switch lenders (Rajan 1992) or, relatedly, for the lenders to sell their portfolio without incurring a substantial discount. In that sense, there is lender-dependence for many borrowers in the middle market, while larger firms tend to be an exception. If a lender becomes constrained, it is a shock that could spillover to the borrowers and the broader economy.

Neither business development companies nor credit funds are runnable, as they generally fund themselves with equity and long-term debt. However, their use of

leverage and reliance on market funding are two potential constraints that warrant careful attention when evaluating the structural vulnerabilities of private creditors.

The vast literature on banking shocks and their spillovers points to the *use of leverage* as a central mechanism in the propagation of financial stress. Private debt lenders operate with significantly more conservative leverage limits and capitalization standards than banks—for example, business development companies are typically subject to a 2:1 debt-to-equity limit, compared to approximately 10:1 for banks. As a result, even relatively small shocks to a bank's asset value can cause capital constraints to become binding, prompting banks to contract lending to meet regulatory requirements. In contrast, due to their lower leverage, business development companies are far less sensitive to such shocks.

Still, loans on the balance sheets of private credit providers are also much riskier than those held by banks, and, therefore, warrant higher capital levels. Using the bank regulatory framework to assess capitalization of business development companies, Chernenko, Ialenti, and Scharfstein (2025) find that as of 2023, the median risk-based capital ratio for business development companies was approximately 36 percent, compared to 10 percent for banks. They also find business development companies to be resilient to bank stress-testing frameworks. Thus, based on existing standards, the current capitalization of private credit lenders exceeds regulatory benchmarks by a significant margin.

An additional point to highlight is that, traditionally, banks have been reluctant to raise equity to recapitalize following negative shocks to their assets. A plausible reason could be that incumbent equity-holders do not wish to dilute their holdings, and would instead prefer to make higher-risk loans in a gamble for resurrection—or hope for a government bailout. However, alternative-asset conglomerates are rich in unused capital. Capital commitments in private equity, private debt, and other types of alternative funds are typically raised—and tend to significantly expand—during benign economic cycles. As a result, an adverse macroeconomic shock often finds alternative funds with substantial "dry powder" capital ready to deploy. Thus, alternative conglomerates have the ability to inject new capital if necessary, rather than enter into portfolio unwinding mode as banks typically do.

Another potential concern for private debt is its *reliance on market funding*. Specifically, business development companies depend on demand for bonds backed by a pool of risky loans. This is similar to pressures faced by the collateralized loan obligations—the key type of institutional investors in the leveraged loan market—which, as shown in Figure 4, can fluctuate substantially in response to market conditions. From a financial stability perspective, credit fund structures are less sensitive to market fluctuations. Indeed, all major private debt managers employ a range of alternative funding structures, which—at least anecdotally—they view as complementary tools to diversify funding sources.

Exposures of Banks to Private Creditors

Roughly 20 percent of funding of business development companies comes from bank-provided revolving credit lines. Several studies—including Berrospide,

Cai, Lewis-Hayre, and Zikes (2025), Fillat, Landoni, Levin, and Wang (2025), and Haque, Jang, and Wang (2025)—have examined this connection. The question is whether this indirect exposure of the banking sector could serve as a channel through which potential problems in private creditor portfolios might spill over into the banking system.

To understand banks' potential exposure, we must consider how risk can build up within a private credit portfolio. Two subtle mechanisms—both with precedents in the history of high-yield debt markets—deserve close attention: (1) a deterioration in credit standards and (2) the use of contingent credit.

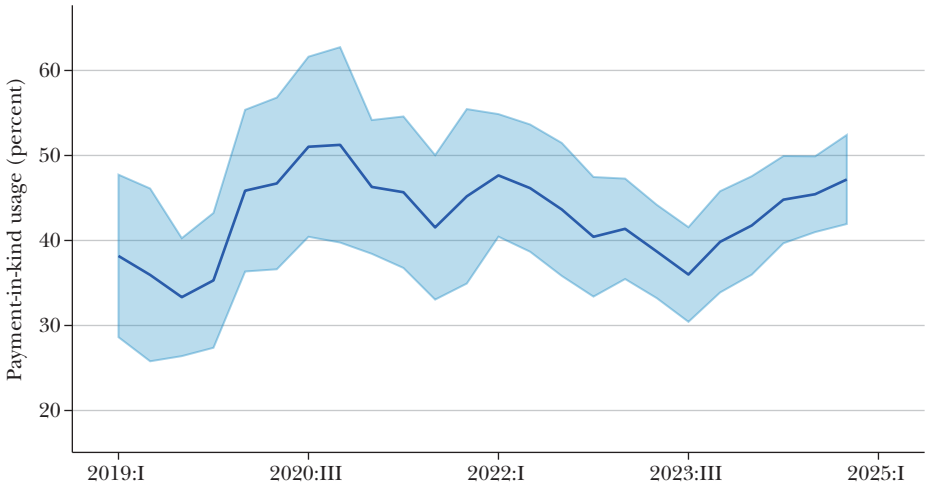
In the first setting, competitive pressures or the drive for rapid growth could push private creditors to loosen underwriting standards to such an extent that, if losses materialize, they are absorbed by senior creditors rather than by the equity tranche. Such an outcome would be unexpected. These loan agreements are expected to define and protect collateral, ensure creditors' seniority in the capital structure, and put in place a system of reliable early warnings through financial covenants. But the recent history of leveraged loan markets provides an example of how these mechanisms can fail.

In 2009, approximately 30 percent of leveraged loans included quarterly compliance requirements for leverage and interest coverage limits. By 2015, over 90 percent of these loans had eliminated the limits for interest coverage, and the leverage test shifted from periodic to conditional verification. At the same time, leverage limits were weakened in various ways: they started to accommodate higher indebtedness, changes to the definitions of financial variables used for covenant testing, and the introduction of broad exemptions (Ivashina and Vallée 2025). When leveraged creditors faced the COVID-19 shock and its aftermath, which led to a wave of restructurings, senior creditors absorbed significant risk. In 2024, the default rate in the leveraged loan market stood at 1.45 percent—well below its historical average—but an additional 3.25 percentage points of companies underwent distress-driven restructurings (as reported by Kakouris 2025).

It is somewhat early to say whether similar dynamics could emerge in the private credit market. Competitive pressure is certainly present in the private debt industry, and fundraising booms create pressure to deploy capital, a dynamic that warrants particular attention as fundraising from retail investors continues to expand. On the positive side, the private credit market remains largely bilateral, illiquid, and held on lenders' balance sheets—unlike the broadly syndicated loan market, which functions more like an auction involving a wide range of smaller, dispersed creditors. Regardless, contractual data are very scarce.

Separately from covenant protection, private debt is expected to provide flexibility, and one such flexible feature is called payment-in-kind. When a payment-in-kind provision is exercised, the interest rate on a private credit loan usually increases by about two percentage points. For example, suppose that in the first quarter of 2025 a borrower owes $2 million in quarterly interest on an 8 percent loan with $100 million in principal. If the borrower elects to use payment-in-kind, there is no cash outlay, but the interest rate increases to

Figure 5
Prevalence of Payment-in-Kind Usage Among Eligible Borrowers



Source: Ivashina (2026).
Note: The figure provides an estimate of payment-in-kind usage, defined as the ratio of reported in-kind interest to its estimated maximum level. The bold line represents the average across business development companies in the PitchBook data, and the shaded bands indicate 95 percent confidence intervals.

10 percent and the principal rises to $102.5 million. If the borrower continues to use payment-in-kind throughout 2025, the principal would grow to $110.4 million by year-end. Borrowers are more likely to utilize payment-in-kind when they are struggling to meet cash obligations, thereby shifting the risk to the lender. Normally, leverage and interest coverage covenants would help to constrain use of payment-in-kind: conversely, the erosion of covenants protection would enable more aggressive use of payment-in-kind.

Again, some history suggests the concerns that can arise with payment-in-kind contracts. Kaplan and Stein (1993) examined the wave of leveraged buyouts during the 1980s. At that time, high-yield bonds were the primary form of risky debt financing, with approximately 60 percent of bonds backing buyouts including a payment-in-kind feature at the end of the decade. Lacking meaningful covenant protection, high-yield bonds became exposed to "gambling for resurrection," as buyout sponsors took increasingly risky bets when valuations came down in the late 1980s.

Measuring the use of payment-in-kind is not straightforward, even for business development companies where detailed portfolio data are available. One challenge is that firms with these provisions can choose to pay some or all of the interest in cash. To simplify, I use a proxy: total payment-in-kind interest reported by business development companies in their quarterly statements, scaled by the maximum

potential payment-in-kind interest. If all firms with payment-in-kind provisions chose to use them in full, this ratio would be equal to one.

Using this measure, Figure 5 shows that the principal-weighted exercised share of loans with payment-in-kind provisions increased from 36 percent in 2019 to 48 percent during the pandemic. This illustrates how payment-in-kind structures can serve as a channel through which macroeconomic shocks are transmitted to the balance sheets of private creditors.

Payment-in-kind usage has recently risen again, returning to pandemic-era levels in the second half of 2024. This may signal a deterioration in the credit quality of the underlying portfolio.[10] Yet, the magnitudes remained modest through 2025.

Overall, because banks typically provide senior secured debt to well-capitalized structures, even if private creditors were to incur significant losses, it remains unlikely that bank lenders would face impairment. That said, the contingent nature of bank exposure is important: as private debt borrowers begin to exercise payment-in-kind provisions, private creditors may become more likely to draw on their revolving bank credit lines. As of the end of 2025, approximately half of these facilities were utilized, implying that banks' exposure could potentially double just as the credit quality of the underlying portfolios begins to deteriorate. However, this view overlooks important mitigating factors, particularly bank debt maturities and covenant protections; at least historically, banks have imposed tight covenant structures to constrain such lending. If so, there may still be meaningful risks for private creditors, as covenants can become binding early and, in the absence of refinancing options, create significant liquidity pressures for private debt funds. In sum, utilization of revolving credit facilities and their terms warrants careful monitoring.

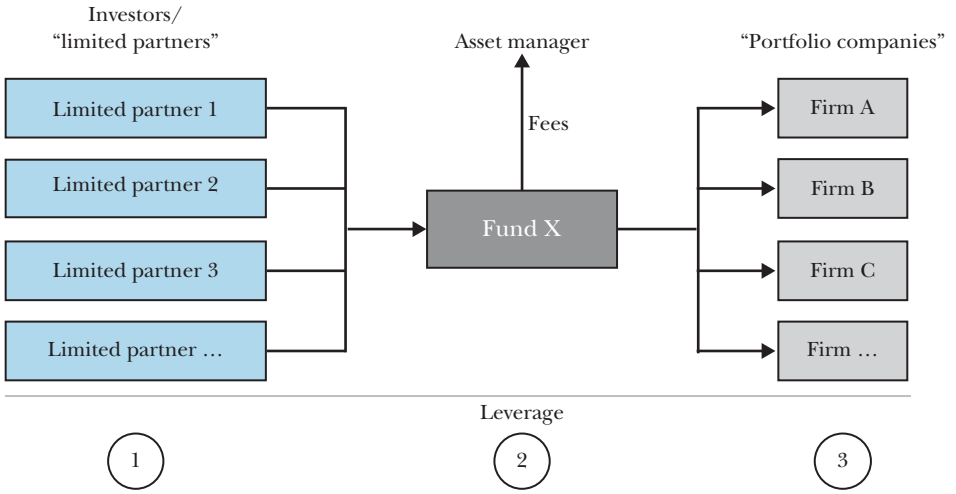
Is There Hidden Leverage?

Another concern raised in policy discussions is whether the expansion of private debt is leading to a worrisome buildup of complexity alongside leverage. Private equity firms have employed leverage to fund buyouts for decades. Today, leverage extends beyond the underlying investment level and is increasingly prevalent at both the fund level (across private equity and private debt) and the fund manager level. This raises concerns that leverage may be additive in periods of stress.

The closest historical parallel is the 2007–2009 global financial crisis, when banks' off-balance-sheet leverage became a major source of stress. In the years leading up to the crisis, large banks commonly moved mortgage-backed securities into off-balance-sheet vehicles funded with short-term debt. In crisis, these assets were forced back onto banks' balance sheets, leading to a sudden recognition of previously embedded leverage.

[10] In additional estimates using a sample of 114 business development companies, I find that income arising from payment-in-kind peaked at 10.1 percent of total investment income in 2020:Q4 during the pandemic, declined to 6.3 percent at the beginning of 2023, and then rose again to 9.2 percent by the end of 2024. Data and an illustrative figure are provided in the Supplemental Appendix available on the JEP website.

Figure 6
Representative Private Equity Fund Structure



Source: Author's creation.
Note: This stylized figure illustrates the intermediation role of private equity managers and highlights three distinct borrowing counterparties that can use private debt.

The rise in the use of leverage throughout the alternative asset management industry warrants close monitoring, but the fact that leverage is employed at both the fund and manager levels is less concerning than it may initially appear. Figure 6 offers a simplified representation of the structure of a typical private equity or private credit fund. The objective is to make investments—debt investments in the case of private debt and equity investments in the case of private equity. This is carried out through a fund owned by investors, also known as "limited partners." The fund is managed by an intermediary, the alternative asset manager, who charges fixed, upfront management fees; these fees are essentially risk-free once the fund is closed.

The numbers below the figure represent the three main points of leverage. Node 1 is the traditional leverage taken on by a portfolio company in a leveraged buyout, issued against expected cash flows of the underlying company and secured by collateral.

Fees to the private equity firm represent a separate and largely independent stream of cash flows from the fund's investment returns. Therefore, both the underlying buyout and the firm managing the fund can have their own capital structures—in other words, both can be leveraged. The most established firms—those managing multiple funds across a range of strategies—are generally better positioned to leverage their franchise value as they have a more stable stream of fees (node 2).

This brings us to leverage at the fund level (node 3). This form of leverage is effectively borne by limited partners and is typically backed either by their uncalled

Table 1
Cross-Market Refinancing Flows: Leveraged Loans and Private Credit

($ billion)	2022	2023	2024	2025:H1
Leveraged loans refinancing private credit loans	2.19	1.99	28.86	16.11
Private credit loans refinancing leveraged loans	10.99	22.37	30.83	12.83
Net refinancing into private credit	8.80	20.38	1.97	−3.28
New issuance of collateralized loan obligations	129.32	115.81	201.94	99.83
Total institutional loan issuance	225.30	234.18	502.17	225.44

Note: Compiled using PitchBook data.

capital commitments or by the invested portfolio. Earlier, I discussed the level of leverage used in private debt fund structures and concluded that, at this stage, it appears relatively conservative. Whether it is appropriate for, say, a pension fund (a representative limited partner) to apply leverage to its portfolio of investments is an important question, but one that lies beyond the scope of this paper.

Finally, today's private debt market is dominated by very large players—such as the "big four" of Blackstone, Apollo, KKR, and Carlyle. Most are alternative asset conglomerates operating across a wide range of investment strategies beyond private debt. Private debt funds represent just one segment of the many investment vehicles they manage, and even individual funds are only one of several structures within broader credit platforms. While this adds further complexity to the analysis, it is also relevant when considering incentives—including reputational capital—as well as institutional resources, sophistication, and access to capital.

Complementarity with High-Yield Corporate Sectors

One final thought about potential macroeconomic spillovers from private debt funds is that the interplay between three distinct but complementary high-yield corporate markets—high-yield bonds, leveraged loans, and private debt—introduces a stabilizing element. For example, during the 2009–2010 period—when issuance of collateralized loan obligations remained shut down (shown in Figure 4)—PitchBook LCD data show that $179 billion in loans were refinanced through high-yield bond issuance.

However, bonds are essentially inaccessible for smaller companies; moreover, as interest rates started to rise after the pandemic, companies looked for flexibility if debt restructuring might become necessary. In this context, private debt became a more attractive option. Table 1 summarizes the total volume of loans refinanced in the leveraged loan market using private debt, and vice versa. The difference between total institutional capital and issuance collateralized loan obligations reflects inflows into loan mutual funds. For example, in 2023, the market for collateralized loan obligations was depressed and loan mutual funds experienced outflows. In this context, $22.4 billion in loans were refinanced by taking on new loans from private credit providers to repay existing leveraged loans. By 2024, the situation began to

shift as the leveraged loan market recovered, with the balance slightly reversing back toward the leveraged loans market in the first half of the year.

While switching lenders likely remains prohibitively costly for smaller companies, the coexistence of multiple sources of high-yield capital offers larger borrowers valuable funding diversification, which can be beneficial not only for the companies themselves but ultimately for the broader economy.

To conclude, my overall assessment is that, as of the first quarter of 2026, the systemic risk posed by the expansion of private debt remains moderate. However, private debt continues to grow and evolve. We may be approaching a critical juncture, as the anticipated entry of 401(k) plans into private assets could bring a substantial amount of new capital in search of investment opportunities. At the same time, the easing of supervisory and regulatory requirements for banks could lead to further increases in leverage. Continuous engagement among policymakers, industry participants, and academics will be essential to understanding the potential macroeconomic implications.

■ *I thank Amy Rose and Zach Wallace-Wright for their exceptional assistance in preparing this paper. I am also very grateful for the detailed and helpful comments from the editors, as well as Falk Bräuning, Paul Gordon, Steven Kaplan, Daniel Tarullo, David Scharfstein, Dwight Scott, Adi Sunderam, and Bill Vrattos.*

References

Acharya, Viral V., Nicola Cetorelli, and Bruce Tuckman. 2026. "Where Do Banks End and NBFIs Begin?" *Review of Corporate Finance Studies*. https://doi.org/10.1093/rcfs/cfag012.

Avalos, Fernando, Sebastian Doerr, and Gabor Pinter. 2025. "The Global Drivers of Private Credit." *BIS Quarterly Review* (March 2025): 13–30.

Bain & Company. 2024. *Global Private Equity Report 2024*. Bain & Company. Inc.

Berrospide, Jose, Fang Cai, Siddhartha Lewis-Hayre, and Filip Zikes. 2025. "Bank Lending to Private Credit: Size, Characteristics, and Financial Stability Implications." FEDS Notes, May 23. https://www.federalreserve.gov/econres/notes/feds-notes/bank-lending-to-private-credit-size-characteristics-and-financial-stability-implications-20250523.html.

Block, Joern, Young Soo Jang, Steven N. Kaplan, and Anna Schulze. 2024. "A Survey of Private Debt Funds." *Review of Corporate Finance Studies* 13 (2): 335–83.

Bolton, Patrick, and David S. Scharfstein. 1996. "Optimal Debt Structure and the Number of Creditors." *Journal of Political Economy* 104 (1): 1–25.

Chernenko, Sergey, Isil Erel, and Robert Prilmeier. 2022. "Why Do Firms Borrow Directly from Nonbanks?" *Review of Financial Studies* 33 (11): 4902–47.

Chernenko, Sergey, Robert Ialenti, and David S. Scharfstein. 2025. "Bank Capital and the Growth of Private Credit." Preprint, SSRN. https://dx.doi.org/10.2139/ssrn.5097437.

Cortes, Fabio, Mohamed Diaby, Caio Ferreira, Nila Khanolkar, Harrison Samuel Kraus, Benjamin Mosk, Natalia Novikova, Nobuyasu Sugimoto, Dmitry Yakovlev, and Charles Cohen. 2024. "The Rise and

Risks of Private Credit." In *Global Financial Stability Report, The Last Mile: Financial Vulnerabilities and Risks*, 53–76. International Monetary Fund.

Davydiuk, Tetiana, Tatyana Marchuk, and Samuel Rosen. 2024. "Direct Lenders in the U.S. Middle Market." *Journal of Financial Economics* 162: 103946.

Federal Reserve Bank of New York. 2026. "Effective Federal Funds Rate [EFFR]." Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/EFFR (accessed April 2026).

Fillat, José L., Mattia Landoni, John D. Levin, and J. Christina Wang. 2025. "Could the Growth of Private Credit Pose a Risk to Financial System Stability?" Federal Reserve Bank of Boston Current Policy Perspectives Paper 25–8.

Financial Policy Committee. 2025. *Financial Stability Report—July 2025*. Bank of England.

Haque, Sharjil, Young Soo Jang, and Jessie Jiaxu Wang. 2025. "Indirect Credit Supply: How Bank Lending to Private Credit Shapes Monetary Policy Transmission." FEDS Working Paper 2025-059.

Ivashina, Victoria. 2026. *Data and Code for: "The Role of Private Debt in the Financial Ecosystem."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E248034V1.

Ivashina, Victoria, and Josh Lerner. 2018. "Looking for Alternatives: Pension Investments around the World, 2008 to 2017." Unpublished.

Ivashina, Victoria, and Srimayi Mylavarapu. 2025. "Arsenal Capital Partners' Refinancing of Pinnacle." Harvard Business School Case 225–044.

Ivashina, Victoria, and Boris Vallée. 2025. "Weak Credit Covenants." *Management Science* 71 (12): 9998–10021.

Kakouris, Rachelle. 2025. "Data Dive: Leveraged Loan Default Stats Show Rift in Restructuring Landscape." *PitchBook News and Analysis*, January 30. https://pitchbook.com/news/articles/data-dive-leverage-loan-default-stats-show-rift-in-restructuring-landscape.

Kaplan, Steven N., and Jeremy C. Stein. 1993. "The Evolution of Buyout Pricing and Financial Structure in the 1980s." *Quarterly Journal of Economics* 108 (2): 313–57.

Kaplan, Steven N., and Per Strömberg. 2009. "Leveraged Buyouts and Private Equity." *Journal of Economic Perspectives* 23 (1): 121–46.

Kashyap, Anil K., Raghuram Rajan, and Jeremy C. Stein. 2002. "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking." *Journal of Finance* 57 (1): 33–73.

PitchBook Datafeed. n.d. "Venture Capital and Private Equity Universes." Harvard Business School. (Accessed July 2025.)

PitchBook LCD. 2026a. "US Leveraged Loan Quarterly Trend Lines—Q1 2026". Harvard Business School. https://pitchbook.com (accessed May 2026).

Pitchbook LCD. 2026b. *BDC Portfolio Holdings, 2004Q1-2025Q1*. Harvard Business School https://pitchbook.com (accessed May 2026).

Prequin Ltd. n.d. "Prequin Pro." Harvard Business School. https://pro.prequin.com/ (accessed May 2026).

US Securities and Exchange Commission. 2012–2026. "Business Development Company Report, 2012–2025." US Securities and Exchange Commission. https://www.sec.gov/data-research/sec-markets-data/opendatasetsshtmlbdc (accessed April 2026).

US Securities and Exchange Commission. 2012–2026. "EDGAR filings and forms." US Securities and Exchange Commission. https://www.sec.gov/edgar/search/ (accessed October 2025).

Standard and Poors. 2026. "Compustat North American Fundamentals Annual Data." Wharton Research Data Services (WRDS). https://wrds-www.wharton.upenn.edu/login/?next=/pages/get-data/compustat-capital-iq-standard-poors/compustat/north-america-daily/fundamentals-annual/ (accessed May 2026).

Rajan, Raghuram G. 1992. "Insiders and Outsiders: The Choice between Informed and Arm's-Length Debt." *Journal of Finance* 47 (4): 1367–400.

Sundar, Sitara. 2025. "Why Private Credit Remains a Strong Opportunity." JP Morgan Private Bank, July 31. https://privatebank.jpmorgan.com/nam/en/insights/markets-and-investing/why-private-credit-remains-a-strong-opportunity.

Journal of Economic Perspectives—Volume 40, Number 4—Fall 2026—Pages 95–114

Private Credit, Balance Sheets, and Financial Stability

Gregor Matvos, Tomasz Piskorski, and Amit Seru

Private credit has grown rapidly over the past two decades and has become an increasingly important source of financing for businesses. More broadly, the growth of private credit is part of a secular shift in financial intermediation away from traditional banking and toward nonbank intermediaries, which now fund the majority of credit to US households and firms (Buchak, Matvos, Piskorski, and Seru 2024c).[1] Policymakers, regulators, and market participants increasingly view private credit through the lens of "shadow banking," raising concerns that it may recreate vulnerabilities traditionally associated with banking—including high leverage, opacity, maturity mismatch, and interconnectedness—outside the regulatory perimeter. Recent redemption pressures in some private credit

■ *Gregor Matvos is Professor of Finance, Northwestern Kellogg School of Business, Evanston, Illinois. Tomasz Piskorski is Professor of Finance and Real Estate, Columbia Business School, New York City, New York. Amit Seru is Professor of Finance, Stanford Business School, as well as Senior Fellow, Hoover Institution, and Senior Fellow, Stanford Institute for Economic Policy Research, all at Stanford, California. The authors are listed in alphabetical order. All three authors are also Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are gregor.matvos@kellogg.northwestern.edu, tp2252@gsb.columbia.edu, and aseru@stanford.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20261498.

[1] Recent research has examined the causes and consequences of this shift toward nonbank intermediation, including the roles of regulation, technology, monetary policy, investor demand, and funding linkages between banks and nonbanks (Buchak, Matvos, Piskorski, and Seru 2018, 2024a–c; Jiang, Matvos, Piskorski, and Seru 2024b; Xiao 2020; Irani, Iyer, Meisenzahl, and Peydró 2021; Hachem and Song 2021; Begenau and Landvoigt 2022; Cherry et al. 2022; Gopal and Schnabl 2022; Chernenko, Erel, and Prilmeier 2022; Jiang 2023; Acharya, Gopal, Jager, and Steffen 2025).

vehicles have further intensified these concerns and renewed questions about the implications of private credit for financial stability.

These concerns reflect a broader lesson from the history of financial crises. Financial instability rarely arises simply because asset values decline. Rather, crises emerge when losses interact with fragile funding structures. Traditional banks provide the classic example. Banks hold long-term and often illiquid assets while funding themselves primarily with short-term liabilities that can be withdrawn on demand. When confidence deteriorates, this combination of leverage, maturity mismatch, and runnable liabilities can transform losses into systemic crises.

From this perspective, the key issue for financial stability is not who provides credit but how that credit is funded. The central question is therefore whether private credit funds replicate the balance-sheet structures that make banks fragile. Do private credit funds rely on substantial leverage and short-term funding? Are they exposed to significant maturity mismatch? How are losses allocated between investors and creditors? To what extent could stress in private credit spill over to the broader financial system?

This article examines these questions using comprehensive fund- and asset-level data on private credit funds, the dominant institutional form of private credit. The data cover approximately 1,300 funds between 2000 and 2024, representing the majority of institutional private credit fund assets in the United States. These data allow us to characterize the capitalization, funding structure, maturity profile, portfolio composition, and performance of private credit funds and to compare these features directly with those of traditional banks.

Four broad facts emerge from the data. First, private credit funds are financed primarily with equity capital provided by long-term investors, with equity typically accounting for 65–80 percent of fund assets, compared with about 10 percent for banks. Second, private credit funds engage in little maturity transformation, with fund lives generally spanning eight to twelve years, while the loans they hold typically mature within two to four years. Third, private credit portfolios are diversified across industries, geographies, and credit strategies. Fourth, equity investors have earned average net annualized returns of about 10 percent after fees, while the substantial equity cushions maintained by private credit funds have historically absorbed losses before debt creditors became exposed.

Taken together, these facts suggest that private credit funds differ in important ways from traditional deposit-funded banks. While private credit funds clearly bear credit risk, they generally do not combine that risk with the funding structures that have historically generated systemic fragility in banking. Nevertheless, important questions remain regarding valuation practices, investor behavior, and the expansion of private credit vehicles that permit investor redemptions. These questions have attracted increasing attention in recent years as some private credit vehicles have experienced redemption pressures despite maintaining relatively conservative balance-sheet structures. We return to them in the concluding section.

The remainder of the paper proceeds as follows. The next section provides institutional background on private credit and describes the data. We then present

evidence on the capitalization, maturity structure, portfolio composition, and performance of private credit funds and discuss the implications of these facts for financial stability. The final section discusses several emerging issues and areas of uncertainty as the sector continues to evolve.

Institutional Background and Data

Private Credit and Its Institutional Forms

Private credit refers to financing provided by specialized investment vehicles to businesses, primarily through loans and other private credit instruments. Unlike traditional bank lending, these investments are funded primarily through capital commitments from investors, rather than deposits. Although private credit remains a relatively small component of the more than $40 trillion of outstanding credit extended to US households and firms, it has expanded rapidly and accounted for more than $1.1 trillion in assets by mid-2024. The sector plays a significant role in financing middle-market firms, where private credit funds have become an increasingly important alternative to traditional bank lending.

Recent research provides insight into the types of firms served by private credit lenders and the lending technologies they employ. Jang, Kim, Sufi, and Chen (2026) show that direct lenders frequently finance private-equity-backed firms and are particularly active in industries characterized by intangible capital and relationship-based lending. Survey evidence in Block, Jang, Kaplan, and Schulze (2024) further suggests that private credit lenders rely heavily on screening, monitoring, and covenant protections when lending to opaque middle-market firms.

The private credit market is organized through several institutional forms. The largest segment consists of private credit funds, which are the focus of this article. By mid-2024, these funds accounted for approximately $750 billion of invested capital and more than $1 trillion of total committed capital when "dry powder" (capital raised by a fund but not yet loaned) is included. Another important segment consists of business development companies (BDCs), which accounted for approximately $375 billion in assets, including roughly $220 billion in nontraded BDCs and the remainder in publicly traded BDCs.[2]

These institutional forms differ importantly in their legal structures and funding models. Private credit funds are typically organized as closed-end investment partnerships financed by long-term capital commitments from institutional investors. These funds have a finite life and return capital to investors as underlying investments are realized. As a result, investors generally cannot redeem their interests during the life of the fund.

Business development companies, by contrast, are investment companies that provide credit to businesses and are structured to allow participation by both

[2]Based on the calculations by Federal Reserve researchers using Preqin and disclosures by business development companies (Berrospide, Cai, Lewis-Hayre, and Zikes 2025).

institutional and retail investors. Publicly traded BDCs face little direct redemption pressure, because investors can exit by selling shares in secondary markets without requiring the vehicle itself to liquidate assets or return capital. Nontraded BDCs, however, typically offer investors periodic liquidity through repurchase programs and have therefore been the primary focus of recent concerns regarding redemption pressures in private credit. Even in this segment, investor withdrawals are generally subject to substantial restrictions, including limits on the amount of capital that investors can request to withdraw during a given period and provisions allowing managers to defer or suspend withdrawals during periods of stress.

These distinctions become important when interpreting recent concerns about liquidity risk and the potential for investor runs in private credit. Recent redemption pressures have been concentrated in nontraded business development companies and other vehicles that permit investors to redeem shares or request repurchases of their investments. While these vehicles represent an important part of the private credit ecosystem, the majority of private credit assets are held by private credit funds organized as closed-end vehicles. Because private credit funds represent the largest segment of the private credit market, they are the focus of our analysis. We therefore begin by describing how these funds are organized and financed.

Private Credit Funds: The Dominant Institutional Form of Private Credit

Private credit funds are typically structured as limited partnerships (Block, Jang, Kaplan, and Schulze 2024). A fund manager, known as the general partner, is responsible for sourcing, underwriting, and monitoring investments, while outside investors, known as limited partners, provide most of the capital. In practice, general partners are often specialized private credit firms or large alternative asset managers. Limited partners are primarily institutional investors such as insurance companies, pension funds, endowments, foundations, and sovereign wealth funds that allocate capital to private credit in search of yield, diversification, and long-term returns. The general partner typically contributes a small amount of capital alongside investors and is compensated through a combination of management fees and performance-based carried interest.

These funds primarily invest in loans and other credit instruments extended to businesses, often targeting middle-market firms that have limited access to public debt markets or traditional bank financing. The sector encompasses a range of lending strategies, including senior lending, mezzanine lending, distressed debt investing, and generalist credit funds. Senior lending focuses on loans that are highest in repayment priority, mezzanine lending combines debt-like and equity-like features, distressed debt strategies target financially troubled firms, and generalist funds invest across multiple segments of the credit market. Most funds have a defined investment period and finite life, during which capital is called from investors, deployed into loans and other investments, and ultimately returned as investments are realized. Investors generally do not have the ability to withdraw capital at their discretion during the life of the fund.

Data and Sample Coverage

Our analysis uses data from the MSCI Private Capital Universe, one of the most comprehensive sources of information on institutional private credit funds. The database includes fund-level observations for 1,267 private credit funds primarily active in the United States. The sample spans 2000–2024 and includes both active and fully exited funds. The data are assembled from information reported by fund managers and institutional investors participating in these funds and are used by investors to monitor fund holdings, valuations, and performance.

In our sample, the number of private credit funds increased from fewer than 30 funds in 2000 to more than 900 by 2023, illustrating the rapid expansion of private credit as an institutional asset class. There was a corresponding increase in assets invested by private credit funds, from less than $10 billion in the early 2000s to more than $600 billion by the end of the sample period. By mid-2024, invested capital in private credit funds exceeded $750 billion, while total committed capital exceeded $1 trillion when dry powder is included. Comparing Federal Reserve estimates based on private-sector data from Preqin with the assets captured in the MSCI sample suggests that coverage improves substantially over time. In recent years, the MSCI data capture approximately 80–90 percent of invested private credit fund assets, indicating broad coverage of the dominant institutional segment of private credit.[3]

Each fund observation contains information on total assets, liabilities, credit line borrowing, investment strategies, and performance. The dataset also includes annualized net returns to investors and information on the timing of both fund life cycles and individual investments. In addition, the dataset includes information on 8,982 underlying investments, allowing us to examine the composition of fund portfolios across industries, geographies, and credit strategies, as well as the maturities of individual investments. Together, these data allow us to assess capitalization, funding structure, portfolio diversification, and maturity transformation.

It is important to note that not all private credit funds in the dataset report information on usage of bank credit lines. A missing value does not necessarily imply nonusage and may instead reflect incomplete reporting. Because credit lines represent an important source of debt financing for nonbank lenders and are primarily provided by commercial banks (Jiang, Matvos, Piskorski, and Seru 2024b; Jiang 2023; Buchak, Matvos, Piskorski, and Seru 2024c; Acharya, Gopal, Jager, and Steffen 2025), understanding their use is important for assessing leverage and bank–nonbank linkages. To address this issue, we also consider a subsample of 574 funds (approximately 45 percent of the sample) that explicitly report positive credit line usage. This subsample allows for a more precise assessment of leverage, funding patterns, and the role of bank-provided credit lines in private credit intermediation.

All results presented below are descriptive. We focus on means, medians, and distributional measures of fund characteristics. Our primary measure of

[3] Supplemental Appendix Figure A1 provides additional details on the growth of private credit and the coverage of our sample.

capitalization is the equity-to-assets ratio, which captures the fraction of assets financed by equity rather than debt. Where appropriate, we report both equal-weighted and asset-weighted statistics to distinguish between the characteristics of the typical fund and those of the largest funds in the market.

Private Credit and Financial Stability: Four Facts

Concerns about the financial stability implications of private credit largely revolve around a familiar question: Do private credit funds replicate the balance-sheet structures that have historically made banks fragile? Research on banking crises and financial instability has emphasized several mechanisms through which financial intermediaries can amplify shocks, including high leverage, maturity transformation, concentrated exposures, and funding structures that transmit losses to creditors or trigger investor runs. To assess whether similar concerns arise in private credit, we examine four basic characteristics of private credit funds: capitalization, maturity structure, diversification, and the allocation of losses.

This balance-sheet perspective has deep roots in economic theory. Models beginning with Diamond and Dybvig (1983) show that systemic fragility arises not only from risky assets themselves, but also from fragile financing structures that combine thin equity, maturity mismatch, and runnable funding. Recent work confirms that these mechanisms remain central to modern banking crises, including episodes driven by uninsured deposits and interest-rate risk (Egan et al. 2017; Jiang, Matvos, Piskorski, and Seru 2024a, 2024b, 2025; Dávila and Goldstein 2023; Amador and Bianchi 2024). In this issue, Ivashina reviews the emerging literature on private credit and highlights several potential channels through which it could affect financial stability. However, systematic evidence on the balance sheets and performance of private credit funds remains limited. We therefore begin by examining the capitalization of private credit funds.

Fact 1: Private Credit Funds Are Primarily Financed with Equity Capital

Capitalization plays a central role in most discussions of financial stability because leverage amplifies losses and increases the likelihood that adverse asset shocks threaten the solvency of financial intermediaries. Financial crises often arise when relatively small losses are imposed on institutions funded primarily with debt. We therefore begin by examining how private credit funds are financed.

The equity-to-assets (E/A) ratio measures the proportion of a fund's assets financed with equity rather than debt. It therefore provides a direct indicator of leverage and the size of the loss-absorbing buffer protecting creditors. Holding asset risk constant, higher equity capitalization implies lower leverage and greater protection for lenders.

Table 1 provides clear evidence that private credit funds maintain high equity capitalization. Across all funds, the average equity-to-assets ratio equals 85 percent and the median equals 98 percent, indicating that most funds are overwhelmingly

Table 1
Private Credit Debt Funds Capitalization Versus Traditional Banks

	Mean	1%	10%	25%	Median	75%	90%	Observations
Panel A. All private credit funds								
Fund total assets (in millions of dollars)	552.1	0.4	12.2	62.0	218.8	610.0	1,429.4	9,355
E/A ratio (in percent)	85.1%	17.4%	45.6%	78.2%	98.1%	99.8%	100%	9,355
E/A ratio (asset weighted, in percent)	79.6%							9,355
Panel B. Private credit funds with positive credit line usage								
Fund total assets (in millions of dollars)	905.6	15.3	88.8	228.1	517.8	1,109.2	2,137.0	2,250
E/A ratio (in percent)	65.9%	17.4%	27.5%	46.4%	70.2%	88.9%	96.0%	2,250
E/A ratio (asset weighted, in percent)	65.9%							2,250
Panel C. Traditional banks								
Total assets (in millions of dollars)	3,185.4	16.8	51.6	99.0	212.9	519.5	1,443.0	85,941
E/A ratio (in percent)	11.5%	3.9%	7.8%	9.0%	10.4%	12.4%	15.2%	85,941
E/A ratio (asset weighted, in percent)	10.6%							85,941

Source: MSCI Private Capital Universe and US Call Reports.
Note: This table reports the mean equity-to-assets ratio in percentage terms—calculated as total equity capital divided by total assets—for US private credit debt funds from 2000–2024 period, along with the distribution of this ratio across funds. We calculate the E/A ratio at the fund-year level. The sample comprises 1,267 private credit debt funds (panel A), of which 574 funds (approximately 45 percent) report credit line usage. These credit-line-using funds are concentrated in the 2010–2024 period and constitute the sample reported in panel B. Panel C presents comparable statistics for all US banks over the same period (2010–2024), based on year-end Call Report data. On average, private credit funds are more than six times better capitalized than US banks, reflecting their substantially higher E/A ratios.

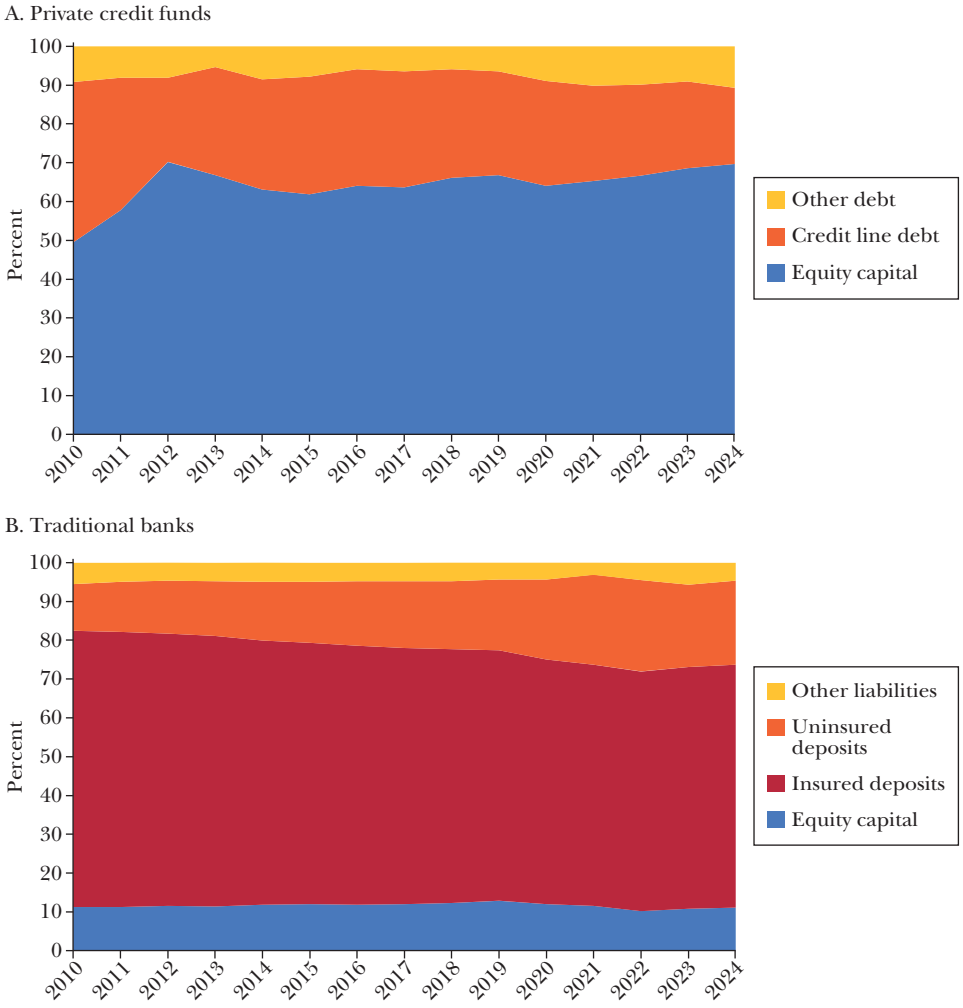
financed with investor equity. Even at the 25th percentile of the distribution, the E/A ratio remains 78 percent, implying limited reliance on borrowing. Asset-weighted statistics show a similar pattern. The asset-weighted average E/A ratio equals approximately 80 percent, indicating that capitalization remains high even among the largest funds.

Substantial capitalization persists also among funds that report positive credit line usage. These credit lines are typically provided by commercial banks and are often used to bridge capital calls, manage transaction timing, and smooth cash flows between investments and investor contributions. Among these funds, both the average and asset-weighted equity-to-assets ratios are about 65 percent. Thus, even when credit lines are used, roughly two-thirds of assets remain financed with equity capital, indicating that credit lines supplement rather than replace equity financing.

Panel C of Table 1 contrasts these patterns with the capitalization of US commercial banks. The average equity-to-assets ratio for banks is approximately 11.5 percent, while the corresponding asset-weighted ratio is about 10.6 percent. Private credit funds therefore operate with equity cushions that are more than six times larger than those of traditional banks.

Figure 1 places these differences in a time-series context. Focusing on funds that report positive credit line usage, the figure shows that capitalization remains relatively stable despite rapid growth in assets. Equity consistently accounts for roughly two-thirds of assets, with credit line borrowing representing most remaining

Figure 1
Capital Structure of Private Credit Funds Over Time



Source: MSCI Private Capital Universe and US Call Reports.
Note: Panel A shows the evolution of the mean liability structure of 574 US private credit funds that report credit line usage over the 2010–2024 period. The stacked area chart decomposes total liabilities into equity capital, credit line debt, and other debt, expressed as percentages of total assets. Panel B presents the corresponding results for US banks, where bank liabilities are decomposed into equity capital, insured deposits, uninsured deposits, and other liabilities, also as shares of total assets. On average, private credit funds are more than six times better capitalized than US banks, reflecting their substantially higher equity-to-assets ratios.

liabilities. Banks exhibit a markedly different structure. Deposits—both insured and uninsured—constitute the majority of bank liabilities, while equity remains a small fraction of total assets. As recent research shows, the growing share of uninsured

deposits has increased banks' exposure to withdrawal risk (Jiang, Matvos, Piskorski, and Seru 2024a, 2025; Granja et al. 2025).

The strong capitalization of private credit funds is not limited to a particular subset of the market. A deeper dive into our sample shows that high equity capitalization persists across lending strategies. Generalist, mezzanine, and distressed funds consistently finance the majority of assets with equity capital, typically maintaining equity ratios between roughly 65 and 80 percent. Senior funds exhibit somewhat lower capitalization and greater use of credit line financing, but even in this segment, equity accounts for the majority of assets. To the extent that mezzanine and distressed strategies are exposed to greater credit risk, their relatively high equity capitalization is economically intuitive because larger equity cushions provide greater protection to creditors against potential losses.

High equity capitalization also persists across industries. Funds lending to general corporate borrowers, real estate firms, and financial firms all finance the majority of assets with equity capital throughout the sample period. Although the composition of liabilities varies somewhat across sectors, substantial equity cushions are maintained in each case, providing considerable protection to creditors against potential losses. This pattern suggests that strong capitalization is a broad feature of private credit intermediation, rather than a characteristic of a particular lending strategy or industry segment.[4]

These findings are consistent with Jiang, Matvos, Piskorski, and Seru (2024b), who show that nonbank mortgage lenders rely primarily on equity capital and bank credit lines rather than deposits, and consequently maintain much lower leverage than banks. In a calibrated equilibrium model, they attribute banks' substantially higher leverage relative to nonbank lenders to their access to insured deposit funding and their ability to issue money-like claims. Our results indicate that this pattern extends beyond mortgage lending to the broader private credit fund sector. The evidence is also consistent with Block, Jang, Kaplan, and Schulze (2024), who report limited leverage among private debt funds. In addition, our calculations based on public disclosures of business development companies—another important segment of private credit—indicate that they have average equity-to-assets ratios of roughly 50 percent, substantially above those observed for traditional banks. Similarly, Chernenko, Ialenti, and Scharfstein (2025) document that BDCs operate with relatively low leverage.

Overall, the evidence indicates that private credit funds rely primarily on equity financing rather than debt funding and maintain substantially higher levels of capitalization than traditional banks, even during a period of exceptionally rapid growth.

[4]For additional details on private credit funds across lending strategies, see Supplemental Appendix Figure A.2. For additional details on fund lending across industries, see Supplemental Appendix Figure A.3.

Fact 2: Private Credit Funds Engage in Little Maturity Transformation

A central concern in the banking literature is maturity transformation: the financing of long-duration assets with short-term liabilities. Such structures can make financial intermediaries vulnerable to liquidity pressures and runs when creditors seek repayment before underlying assets mature. Diamond and Dybvig (1983) emphasize how maturity transformation can generate fragility when institutions fund long-term investments with liabilities that may be withdrawn before the underlying assets are realized. Subsequent work shows that these vulnerabilities remain central to modern banking crises, particularly when institutions rely heavily on uninsured deposits and are exposed to credit or interest-rate risk (Egan et al. 2017; Jiang, Matvos, Piskorski, and Seru 2024a, 2024b, 2025; Dávila and Goldstein 2023; Amador and Bianchi 2024). Motivated by this literature, we examine the relationship between fund life cycles and the maturities of the investments held by private credit funds.

In our sample, the average duration for exited private credit funds is approximately 11.6 years, with a median duration of 11.6 years. Even among funds that report bank credit line usage, average duration remains approximately 8.3 years and median duration is approximately 8.1 years. Overall, private credit funds are typically long-lived investment vehicles.[5]

Figure 2 relates the *maturity* of underlying investments to fund duration. Two features stand out. First, funds with longer life cycles tend to hold investments with longer maturities, suggesting that managers broadly align asset maturity with fund duration. Second, investment maturity remains substantially shorter than fund duration across the entire distribution. For example, funds with durations of approximately 8 years hold investments with average maturities of only about 2.5 years, while funds with durations of approximately 12 years hold investments with average maturities of roughly 4 years. Thus, investments are typically realized several years before investors' capital must be returned.

This structure differs markedly from that of traditional banks. Because private credit funds are organized as closed-end vehicles and generally do not permit early investor withdrawals, fund liabilities typically extend beyond the maturity of underlying investments. Rather than financing long-duration assets with short-duration liabilities, private credit funds generally hold assets that mature well before capital must be returned to investors.

Overall, the evidence indicates that private credit funds engage in little maturity transformation. Fund life cycles generally exceed the maturities of underlying investments by a substantial margin. For representative funds in our sample, fund durations are typically on the order of eight to twelve years, while underlying investments generally mature within two to four years. As a result, private credit funds do not replicate one of the central balance-sheet characteristics associated with traditional banking fragility.

[5] Supplemental Appendix Table A.1 reports more detailed data on fund duration.

Figure 2
Investment Maturity and Fund Life Duration (in Years)



Source: MSCI Private Capital Universe.
Note: This figure plots the mean and median investment maturity (in years) against the fund age for US private credit funds. Fund life duration is measured from fund inception to the fund's last observed valuation date, while for active funds it is measured from fund inception to the end of the sample period. Investment maturity is measured using exited holdings, from the holding's first investment date to its exit date.

Fact 3: Private Credit Funds Hold Diversified Portfolios

High capitalization and limited maturity transformation reduce the vulnerability of private credit funds to financial distress. However, financial stability also depends on the composition of asset portfolios. Even well-capitalized intermediaries can become vulnerable if their exposures are concentrated in a small number of industries, borrowers, or regions. We therefore next examine the extent of diversification in private credit portfolios.

Table 2 reports the industry composition of 8,982 underlying investments representing approximately $477 billion in total asset value. The data reveal substantial diversification across sectors. Financials account for the largest share of assets at 20.6 percent, followed by industrials (15.4 percent) and health care (12.4 percent). Information technology, consumer discretionary, and real estate each account for roughly 10 percent of total assets. No single industry accounts for more than one-fifth of portfolio value, and investments are spread across a broad range of sectors including communication services, energy, consumer staples, materials, and utilities.

The distribution of holdings closely mirrors the distribution of asset values, suggesting that diversification reflects a large number of underlying investments rather than a small number of concentrated positions. Thus, while individual investments may be exposed to idiosyncratic credit risk, portfolio exposures are generally spread across a diverse set of borrowers and industries.

Table 2
Asset-Level Data: Composition of Private Credit Holdings

Industry of holding	Percent of total assets	Percent of total holdings
Financials	20.6%	18.9%
Industrials	15.4%	17.7%
Health care	12.4%	11.1%
Information technology	11.1%	11.6%
Consumer discretionary	10.3%	11.4%
Real estate	9.9%	10.2%
Communication services	6.1%	4.7%
Energy	4.4%	3.7%
Consumer staples	3.0%	3.5%
Materials	2.9%	3.5%
Utilities	2.0%	1.8%
Unknown/not classified	1.5%	1.5%
All	100%	100%

Data sources: MSCI Private Capital Universe.
Note: This table reports the industry composition of 8,982 asset-level holdings by private credit funds, with a combined asset value of $477.4 billion. It presents each industry's share of total asset value and total holdings. Financials, industrials, and health care together account for nearly half of total assets. Information technology, real estate, and consumer discretionary also represent significant portions of the aggregate portfolio, each comprising approximately 10 percent of total assets. Overall, the distribution of investments across industries indicates substantial sectoral diversification among US private credit fund holdings.

We also examined fund strategies and industry mandates. Distressed and generalist strategies together account for approximately 60 percent of total assets, while senior and mezzanine funds account for most of the remainder. Similarly, although some funds specialize in sectors such as real estate, financials, or technology, approximately 60 percent of assets are managed by funds with broad industry mandates. These patterns suggest that diversification is a feature not only of individual portfolios but also of the private credit sector as a whole.[6]

Geographic diversification further reduces concentration risk. Private credit investments are distributed across a broad range of US states. Although New York (10.3 percent), Texas (7.5 percent), and California (5.7 percent) account for the largest shares of total investment value, together they represent only about one-quarter of total exposure. The remaining investments are broadly distributed across other states, indicating that private credit portfolios are diversified not only across industries but also across geographic regions.[7]

Taken together, the evidence indicates that private credit portfolios are diversified across industries, investment strategies, borrowers, and geographic regions.

[6]For additional detail on fund strategies, see Supplemental Appendix Table A.2.
[7]For additional state-level detail on investments by private credit funds, see Supplemental Appendix Figure A.4.

While individual investments remain exposed to credit risk, diversification reduces the likelihood that distress in any single industry or region translates into large losses for the sector as a whole. As a result, private credit funds are likely to be most vulnerable to broad aggregate shocks that affect many industries and regions simultaneously rather than to isolated sector-specific disruptions.

Fact 4: Losses Are Primarily Absorbed by Equity Investors

Having documented the balance-sheet structure, maturity profile, and portfolio composition of private credit funds, we now examine realized returns to equity investors. Fund-level returns reflect several interacting forces, including the risk characteristics of underlying investments, the degree of diversification across borrowers and industries, exposure to correlated economic risks, the amount of leverage employed, and the allocation of losses between equity investors and creditors. As a result, the distribution of realized returns provides useful information about how risks are ultimately absorbed within the capital structure.

Table 3 reports the historical distribution of annualized net returns earned by limited partners across 1,267 private credit funds. Returns are generally positive across a broad range of funds. Across the full sample, the average annualized return equals 9.6 percent, while the median return equals 9.1 percent. Because private credit funds typically operate for eight to twelve years, as documented in Fact 2, sustained negative annual returns can imply substantial cumulative losses over the life of a fund. Only a small share of funds—roughly 2 to 5 percent—experience returns sufficiently negative to substantially erode the capital of limited partners over a typical fund life cycle. The lower tail of the distribution is relatively thin: The 5th percentile return is approximately −3.8 percent, while the 1st percentile return equals −18.1 percent. Thus, meaningful losses do occur, but they are concentrated among a relatively small fraction of funds and are borne primarily by equity investors.

Fund performance varies across investment strategies. Distressed and mezzanine funds exhibit greater dispersion in returns, consistent with their focus on financially distressed borrowers and more junior positions in the capital structure. Senior lending strategies, by contrast, display the most stable performance and the smallest downside tail. Nevertheless, average returns remain relatively similar across strategies, generally ranging from approximately 9 to 10 percent annually. When we focus on returns for fully exited funds, where performance reflects realized cash flows rather than interim valuations, the results are broadly similar.[8] Average annualized returns increase slightly to 9.9 percent, while median returns equal 9.4 percent. The distribution of outcomes also remains comparable to that observed in the full sample. The similarity between active and exited funds suggests that the return patterns documented in Table 3 are not primarily driven by interim valuation practices.

These return patterns are informative when viewed alongside the balance-sheet evidence presented earlier. Private credit funds typically finance 65 to 80 percent

[8]For additional details on returns for fully exited funds, see Supplemental Appendix Table A.3.

Table 3
Net Returns of Private Credit Funds

	Mean	1%	2%	5%	10%	25%	Median	75%	90%
All funds (Observations = 1,267)	9.6	−18.1	−11.5	−3.8	1.2	6.1	9.1	12.7	17.6
By strategy									
Distressed (Observations = 316)	9.6	−18.6	−15.4	−4.2	1.5	5.9	9.1	13.3	18.7
Generalist (Observations = 280)	9.4	−17.6	−11.7	−4.4	0.7	6.6	9.2	12.6	17.5
Mezzanine (Observations = 327)	9.2	−15.7	−10.1	−4.3	−0.6	5.7	9.4	12.7	16.6
Senior (Observations = 196)	8.7	−4.9	−1.9	1.8	3.0	6.3	8.5	11.4	14.2
Unknown/not classified (Observations = 148)	11.9	−19.3	−13.0	−4.2	1.2	5.9	9.0	14.5	22.4
By industry									
Financials (Observations = 46)	12.0	−6.9	−3.4	−0.2	2.5	7.0	9.4	13.4	22.3
Generalist (Observations = 730)	9.5	−17.5	−12.1	−3.1	1.9	6.0	9.1	12.6	17.2
Industrials (Observations = 20)	8.3	−17.8	−14.6	−5.1	−1.1	7.6	8.7	14.4	15.9
Real estate (Observations = 246)	8.3	−20.4	−15.1	−4.9	−1.1	5.6	8.6	11.8	16.4
Unknown/not classified (Observations = 152)	11.9	−8.4	−5.1	−0.02	2.4	6.9	10.3	14.4	20.4
Consumer discretionary (Observations = 11)	9.1	−1.2	−0.3	2.6	7.3	8.5	10.5	11.4	11.8
Consumer staples (Observations = 5)	4.2	−10.0	−9.5	−8.0	−5.4	2.1	5.3	11.9	12.0
Energy (Observations = 19)	6.1	−12.8	−11.2	−6.3	−4.3	2.2	8.0	9.9	16.0
Health care (Observations = 21)	10.4	2.7	3.4	5.4	5.6	6.9	9.7	11.3	14.3
Information technology (Observations = 9)	9.1	−9.9	−8.5	−4.5	2.2	8.6	9.3	12.1	18.5

Data sources: MSCI Private Capital Universe.
Note: This table reports summary statistics of annualized net returns to equity holders (limited partners) for US private credit funds. Reported figures include the mean, selected percentiles, and sample size. Results are shown for the full sample ("All Funds") as well as disaggregated by investment strategy (for example, distressed, generalist, mezzanine, senior) and by primary industry classification.

of assets with equity capital. As a result, substantial asset losses can be absorbed within the equity layer of the capital structure before creditors experience impairment. The observed distribution of returns suggests that realized gains and losses are borne primarily by limited partners rather than transmitted to creditors.

The evidence is also consistent with recent work on private credit fund performance. Erel, Flanagan, and Weisbach (2024) conduct a comprehensive

risk-adjusted benchmarking analysis of private debt funds and conclude that observed returns largely compensate investors for systematic risk exposures. Similarly, Brown et al. (2025) find that private debt funds have historically performed well on a risk-adjusted basis. Complementing our evidence for private credit funds, Robinson and Wallskog (2026) show that investors in publicly traded business development companies, another segment of the private credit market, have historically earned strong risk-adjusted returns.

Overall, the evidence indicates that losses in private credit funds are primarily absorbed by equity investors. Combined with the substantial capitalization documented in Fact 1, the limited maturity transformation documented in Fact 2, and the diversification documented in Fact 3, these results suggest that losses are largely contained within limited partners capital rather than transmitted to debt creditors.

Remaining Questions and Potential Vulnerabilities

The evidence presented in this article points to four central features of private credit funds. Private credit funds are highly capitalized, engage in little maturity transformation, maintain diversified portfolios, and absorb losses primarily through equity investors rather than creditors. Taken together, these characteristics distinguish private credit funds from traditional deposit-funded banks and suggest that private credit does not currently replicate the balance-sheet structures that have historically generated systemic fragility in banking.

Nevertheless, the absence of these traditional vulnerabilities does not imply that there are no other potential sources of risk. In the remainder of this section, we discuss four potential areas of vulnerability: liquidity provision and retail participation, information frictions and the risk of valuation contagion, the transmission of losses through investor balance sheets, and potential effects on the supply of credit to firms.

Liquidity Provision, Retail Capital, and Withdrawal Risk

The evidence presented in this article primarily reflects the balance-sheet structure of traditional private credit funds, which remain the dominant form of private credit intermediation. As discussed above, these funds are organized as closed-end vehicles, financed by long-term capital commitments, and exhibit little maturity transformation or exposure to withdrawal risk.

An important question, however, is whether the continued evolution of the private credit market will preserve these characteristics. In recent years, private credit has increasingly expanded through vehicles that offer investors some degree of liquidity. The most important examples are the nontraded business development companies, which have grown rapidly and have been the primary focus of recent concerns regarding redemption pressures in private credit. Similar features are increasingly found in "interval funds" and evergreen private credit vehicles, which allow investors to periodically withdraw capital subject to various restrictions.

At the same time, large alternative asset managers have increasingly sought to expand retail participation in private credit, while policymakers have considered initiatives that could broaden access to private markets through retirement and other investment vehicles. These developments have the potential to increase the supply of capital to the sector, but they may also alter its funding structure. To the extent that private credit becomes increasingly funded through vehicles offering liquidity to a broader investor base, the sector may gradually move away from the closed-end model that currently limits withdrawal and run risk.

Recent redemption pressures in nontraded business development companies illustrate these concerns. Nontraded BDCs account for only a minority of private credit assets, and redemption restrictions have thus far limited the need for large-scale asset sales. Nevertheless, these developments demonstrate that withdrawal pressures can emerge when investors are offered liquidity against portfolios of relatively illiquid assets. Such withdrawal pressures are more commonly associated with traditional banking and may become increasingly important if liquidity-providing private credit vehicles continue to expand.

Information Frictions and the Risk of Valuation Contagion

Questions about withdrawal risk are closely connected to questions about information and valuation. The history of credit markets suggests that limited transparency can weaken market discipline and create incentives for asset-quality misrepresentation. Studies of mortgage securitization markets, including Piskorski, Seru, and Witkin (2015) and Griffin and Maturana (2016), document how weak monitoring environments contributed to widespread misreporting of loan characteristics prior to the global financial crisis of 2008–2009, undermining investor confidence as asset values deteriorated. More broadly, Akerlof (1970) emphasizes how informational asymmetries can impair market functioning when asset quality is difficult to verify.

Although private credit differs institutionally from securitized mortgage markets, similar informational frictions may arise because the quality of underlying loans is often difficult for outside investors to observe directly. Moreover, most private credit assets are not continuously traded, and valuations frequently rely on internal models, manager assessments, and information from comparable transactions rather than observable market prices. Our comparison of active and fully exited funds provides some reassurance on this dimension. As discussed in Fact 4, funds with realized cash flows exhibit return distributions similar to those observed for the broader sample, suggesting that the historical performance patterns documented in this article are not primarily an artifact of interim valuation practices. Brown, Gredil, and Kaplan (2019) find that some underperforming private equity fund managers may inflate reported returns during fundraising. Although their findings do not imply widespread misreporting, they demonstrate that opaque valuation environments can create incentives for systematic reporting biases.

Nevertheless, these results are based largely on a period that has not subjected the sector to a severe and prolonged stress episode. Recent enforcement actions

involving valuation procedures and conflicts of interest in private credit markets suggest that such concerns are not merely theoretical.

An additional concern is the risk of *valuation contagion*: the possibility that declines in the market value of one set of credit assets lead investors to mark down the value of similar assets elsewhere, even when there has been no change in their underlying cash flows or credit performance. Stress in more transparent segments of credit markets, such as publicly traded business development companies and other traded private credit vehicles, may lead investors to question the valuation of similar assets held throughout the private credit ecosystem even in the absence of realized credit losses, affecting fundraising, investor flows, and the supply of credit. Such effects may be particularly important in vehicles that offer investors some degree of liquidity.

Finally, credit line facilities could become more important as a source of longer-term funding for private credit during periods of market stress. If asset values decline or capital calls become difficult to meet, funds may rely more heavily on these credit lines, increasing spillover concerns.

Investor Balance Sheets and Loss Transmission

A central implication of the evidence presented in this article is that losses in private credit are largely borne by equity investors rather than creditors. Many of these investors participate in multiple private credit funds and other private market strategies simultaneously. As a result, adverse shocks affecting private credit may generate correlated losses across portfolios and induce broader portfolio adjustments. This channel may be particularly relevant for large institutional investors such as insurance companies, pension funds, endowments, and sovereign wealth funds.

Banks may also be indirectly exposed through their relationships with institutional investors. Many large investors utilize bank-provided liquidity facilities, financing arrangements, and other banking services. To the extent that private credit losses affect investor balance sheets, these effects could propagate indirectly through adjustments in funding, portfolio allocations, or borrowing needs.

Credit Supply During Stress

Private credit has become a significant source of financing for middle-market firms. As a result, adverse shocks that reduce capital flows into the sector may affect the availability of credit to these borrowers. If institutional investors—or, increasingly, retail investors—reduce commitments to private credit funds or withdraw capital from private credit vehicles, fundraising may slow and the amount of capital available for lending may decline. The magnitude of these effects remains uncertain and depends on the extent to which private credit can be substituted by other sources of financing when funding conditions deteriorate. If banks, public debt markets, or other lenders can readily replace private credit, disruptions may have limited consequences. If not, reductions in private credit lending could affect investment, growth, and employment among firms that have become increasingly dependent on this source of capital.

More broadly, the growth of private credit highlights the limits of treating legal or industry labels as the relevant unit of analysis. Private credit funds remain the dominant form of private credit intermediation, and the evidence in this article shows that these funds act largely as long-term investors financed by committed equity capital. At the same time, "private credit" also includes vehicles with different funding structures, investor bases, and liquidity provisions, some of which may resemble banking more closely along certain dimensions. What matters for financial stability is not simply whether credit is provided by a bank or a nonbank, but how the intermediary is funded, what liquidity it promises, and how losses are allocated across balance sheets. This perspective is consistent with Buchak, Matvos, Piskorski, and Seru (2026), who argue for analyzing and regulating financial intermediation based on economic functions rather than legal entity classifications.

■ *We want to thank Greg Buchak and Erica Jiang, our frequent collaborators, for discussions that have significantly shaped our views on this subject. We are grateful to the Private Equity Research Consortium, the Institute for Private Capital, and MSCI for their support and for allowing us to use the MSCI Private Capital Universe database, to Wendy Hu at MSCI for incredible assistance in using these. We also thank Sahil Lalwani who provided outstanding research assistance on this project.*

References

Acharya, Viral V., Manasa Gopal, Maximilian Jager, and Sascha Steffen. 2025. "Shadow Always Touches the Feet: Implications of Bank Credit Lines to Non-Bank Financial Intermediaries." NBER Working Paper 33590.

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (3): 488–500.

Amador, Manuel, and Javier Bianchi. 2024. "Bank Runs, Fragility, and Credit Easing." *American Economic Review* 114 (7): 2073–110.

Begenau, Juliane, and Tim Landvoigt. 2022. "Financial Regulation in a Quantitative Model of the Modern Banking System." *Review of Economic Studies* 89 (4): 1748–84.

Berrospide, Jose, Fang Cai, Siddhartha Lewis-Hayre, and Filip Zikes. 2025. "Bank Lending to Private Credit: Size, Characteristics, and Financial Stability Implications." FEDS Notes, May 23. https://www.federalreserve.gov/econres/notes/feds-notes/bank-lending-to-private-credit-size-characteristics-and-financial-stability-implications-20250523.html?.com.

Block, Joern, Young Soo Jang, Steven N. Kaplan, and Anna Schulze. 2024. "A Survey of Private Debt Funds." *Review of Corporate Finance Studies* 13 (2): 335–83.

Brown, Gregory W., Oleg R. Gredil, and Steven N. Kaplan. 2019. "Do Private Equity Funds Manipulate Reported Returns?" *Journal of Financial Economics* 132 (2): 267–97.

Brown, Gregory, Christian Lundblad, and William Volckmann. 2025. "Risk-Adjusted Performance of Private Funds: What Do We Know?" Unpublished.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2018. "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks." *Journal of Financial Economics* 130 (3): 453–83.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2024a. "Beyond the Balance Sheet Model of Banking: Implications for Bank Regulation and Monetary Policy." *Journal of Political Economy* 132 (2): 616–93.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2024b. "Aggregate Lending and Modern Financial Intermediation: Why Bank Balance Sheet Models are Miscalibrated." In *NBER Macroeconomics Annual*, Vol. 38, edited by Martin Eichenbaum, Erik Hurst, and Valerie Ramey, 239–87. University of Chicago Press.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2024c. "The Secular Decline of Bank Balance Sheet Lending." NBER Working Paper 32176.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. Forthcoming. "Distributed Financial Intermediation." *Journal of Finance: Insights and Perspectives*.

Chernenko, Sergey, Isil Erel, and Robert Prilmeier. 2022. "Why Do Firms Borrow Directly from Nonbanks?" *Review of Financial Studies* 35 (11): 4902–47.

Chernenko, Sergey, Robert Ialenti, and David Scharfstein. 2025. "Bank Capital and the Growth of Private Credit." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.5097437.

Cherry, Susan, Erica Jiang, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2022. "Shadow Bank Distress and Household Debt Relief: Evidence from the CARES Act." *AEA Papers and Proceedings* 112: 509–15.

Dávila, Eduardo, and Itay Goldstein. 2023. "Optimal Deposit Insurance." *Journal of Political Economy* 131 (7): 1676–730.

Diamond, Douglas W., and Philip H. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91 (3): 401–19.

Egan, Mark, Ali Hortaçsu, and Gregor Matvos. 2017. "Deposit Competition and Financial Fragility: Evidence from the US Banking Sector." *American Economic Review* 107 (1): 169–216.

Erel, Isil, Thomas Flanagan, and Michael S. Weisbach. 2024. "Risk-Adjusting the Returns to Private Debt Funds." NBER Working Paper 32278.

Gopal, Manasa, and Philipp Schnabl. 2022. "The Rise of Finance Companies and FinTech Lenders in Small Business Lending." *Review of Financial Studies* 35 (11): 4859–901.

Granja, João, Erica Xuewei Jiang, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2025. "Book Value Risk Management of Banks: Limited Hedging, HTM Accounting, and Rising Interest Rates." NBER Working Paper 32293.

Griffin, John M., and Gonzalo Maturana. 2016. "Who Facilitated Misreporting in Securitized Loans?" *Review of Financial Studies* 29 (2): 384–419.

Hachem, Kinda, and Zheng Song. 2021. "Liquidity Rules and Credit Booms." *Journal of Political Economy* 129 (10): 2721–65.

Irani, Rustom M., Rajkamal Iyer, Ralf R. Meisenzahl, and José-Luis Peydró. 2021. "The Rise of Shadow Banking: Evidence from Capital Regulation." *Review of Financial Studies* 34 (5): 2181–235.

Jang, Young Soo, Dasol Kim, Amir Sufi, and Xiangyu Chen. 2026. "The Lending Technology of Direct Lenders in Private Credit." NBER Working Paper 34500.

Jiang, Erica Xuewei. 2023. "Financing Competitors: Shadow Banks' Funding and Mortgage Market Competition." *Review of Financial Studies* 36 (10): 3861–905.

Jiang, Erica Xuewei, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2024a. "Monetary Tightening and U.S. Bank Fragility in 2023: Mark-to-Market Losses and Uninsured Depositor Runs?" *Journal of Financial Economics* 159: 103899.

Jiang, Erica, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2024b. "Which Banks are (Over) Levered? Insights from Shadow Banks and Uninsured Leverage." NBER Working Paper 26903.

Jiang, Erica Xuewei, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2025. "Monetary Tightening, Commercial Real Estate Distress, and US Bank Fragility." *Journal of Political Economy Macroeconomics* 3 (4): 525–73.

Matvos, Gregor, Tomasz Piskorski, and Amit Seru. 2026. *Data and Code for: "Private Credit, Balance Sheets, and Financial Stability."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. Https://doi.org/10.3886/E251261V1.

Piskorski, Tomasz, Amit Seru, and James Witkin. 2015. "Asset Quality Misrepresentation by Financial Intermediaries: Evidence from the RMBS Market." *Journal of Finance* 70 (6): 2635–78.

Robinson, David T., and Melanie Wallskog. 2026. "The Growth of Private Lending and Retail Access to Alternative Investments." NBER Working Paper 34617.

Xiao, Kairong. 2020. "Monetary Transmission through Shadow Banks." *Review of Financial Studies* 33 (6): 2379–420.

Journal of Economic Perspectives—Volume 40, Number 4—Fall 2026—Pages 115–136

Systemic Cyber Risk: Linking Measurement, Market Incentives, and Policy

Susan Athey, Neil Gandal, and Tyler Moore

All organizations have an incentive to reduce the risk that the key personal, financial, or business secret information held by the firm will be breached. Employees often receive reminders about the risk of "phishing" attacks or other approaches used by hackers to break into the organization. However, cyber hackers do not just attempt to break into one organization at a time; instead, they often seek a more systemic cybersecurity weakness that applies to many organizations. *Product-level systemic cyber risk* occurs when there is (1) concentration among technology suppliers and (2) the products of those suppliers exhibit significant cyber vulnerability. *Firm-level systemic cyber risk* occurs when dominant firms in a sector exhibit significant cyber vulnerability.

Another type of cyber risk comes from interconnected systems; it is typically referred to as *interdependent cyber risk* (Kunreuther and Heal 2003b). Such risk comes from contagion via connected networks. This type of cyber risk is similar to the contagion risk associated with finance, where one failed financial institution can lead to a cascade of failures, or to a public health setting where the contagion effect works through the spread of disease. While such risks are sometimes present in cybersecurity, we focus on a simpler and much more pervasive form of systemic

■ *Susan Athey is Professor of Economics, Graduate School of Business, Stanford University, Stanford, California. Neil Gandal is Emeritus Professor of Economics, Tel Aviv University, Tel Aviv, Israel, and Professor of Cyber Studies, The University of Tulsa, Tulsa, Oklahoma. Tyler Moore is Professor of Cybersecurity, The University of Tulsa, Tulsa, Oklahoma. Gandal is the corresponding author at neil-gandal@utulsa.edu. The author names are listed in alphabetical order.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20261513.

risk in which correlated failures can arise, like when a single piece of widely used vulnerable software results in many firms being attacked at the same time.

As traditionally offline activities like medical care and transportation increasingly adopt digital technologies, the risk of systemic events escalates. The most severe cyber incidents can lead to losses amounting to billions of dollars and also have the potential to result in physical harm, loss of life, or even political turmoil. Moreover, past hackers used artificial intelligence (AI) software as an aid to human efforts, but in the last year or so, AI tools have advanced to the point where they can act as autonomous agents, searching for the same weaknesses across many enterprises. As a result, systemic cyber risk has probably increased considerably. The rise of autonomous AI agents makes the economic analysis in our paper critical for policymakers moving forward.

To set the stage for our discussion, it is helpful to begin with some background on the level of cybersecurity risk. There is no authoritative list of total cyber attacks, nor of the subcategory that would be classified as systemic. However, an array of reports suggest that concerns about cybersecurity are large and growing for private firms, nonprofits, and government. Recent examples include *Global Cybersecurity Outlook 2026* from the World Economic Forum (2026), the *Cyber Security Report 2026* from Check Point Research (2026), and the *Cost of a Data Breach 2026* report from IBM (2026).

Ransomware has become the dominant cybercrime threat to organizations, and it is growing fast. Ransomware is a form of malicious software that encrypts an organization's files or systems, rendering them inaccessible until the victim pays a ransom demand. In recent years, attackers have started threatening to leak stolen data to the public if the ransom is not paid. Attackers often breach a target through unpatched vulnerabilities in information technology products. Ransomware attacks have the potential to be systemic because information technology product markets are very concentrated, and the same vulnerability may be exploited to attack multiple victims.

We do have reasonably good data on the specific measures associated with systemic cyber risk: the number of "ransomware" attacks per year, the number of common vulnerabilities and exposures (CVEs) in software, and the number of known exploited vulnerabilities (KEVs).

The US government's National Vulnerability Database repository hosted by the National Institute of Standards and Technology (NIST) tracks software vulnerabilities and distributes them publicly (NIST 2026). Vulnerabilities are standardized using the common vulnerabilities and exposures system. Each disclosed vulnerability is reviewed and assigned a unique CVE identifier once verified. The National Vulnerability Database enriches each CVE record with severity assessments, affected products, remediation guidance, and references. CVE count is a reasonable proxy for the security of a software product, as insecure software will invariably have more vulnerabilities in it. However, there are downsides to utilizing CVE count directly, because not all vulnerabilities are actively exploited by attackers.

Table 1

Ransomware Attacks, Common Vulnerabilities and Exposures, and Known Exploited Vulnerabilities per Year (2021–2025)

Year	Ransomware attacks	CVEs	KEVs
2021	1,816	20,177	311
2022	3,168	25,000	555
2023	5,412	28,848	187
2024	6,058	39,925	186
2025	7,985	48,154	245

Sources: Sources for the data: For KEVs, www.cisa.gov; for ransomware, Mousqueton (2026); and for CVE, NIST. CISA began to report KEV data in 2021.
Note: Ransomware attacks, common vulnerabilities and exposures, and known exploited vulnerabilities over time.

To that end, the Cybersecurity and Infrastructure Security Agency (CISA) started publishing the known exploited vulnerabilities catalog in 2021 (Cybersecurity and Infrastructure Security Agency 2026). The KEV catalog focuses exclusively on the subset of vulnerabilities published in the National Vulnerabilty Database for which there is confirmed evidence of active exploitation. Hence, products with KEVs can be interpreted as especially risky.

Table 1 reports levels and trends in the number of ransomware attacks, common vulnerabilities and exposures, and known exploited vulnerabilities. The table shows that the number of ransomware attacks has skyrocketed in recent years, quadrupling from 2021 to 2025. Also, the number of CVEs published in 2025 are approximately 2.4 times the number published in 2021. While the number of new KEVs per year has been fairly stable since 2023, nearly 1,500 different vulnerabilities have been exploited since 2021, when the CISA began collecting this data.

In this essay, we will show how the cyber vulnerability can be measured with proxies such as numbers of security flaws in products or fewer observed security controls deployed by firms. We can then build on entity-level vulnerability to define systemic cyber risk. While the risk of catastrophic failures cannot be eliminated, we will point out opportunities to create incentives and institutions to mitigate and reduce systemic cyber risk.

Throughout the discussion, we will apply economic principles to the rising danger of systemic cyber risks. The paper is organized around four central economic lessons and uses the case studies, new data, and policy discussion to develop and support them. First, cybersecurity investment is below the social optimum because buyers have limited ability to assess the security quality of software and hardware before making purchases, switching costs are substantial, and software providers often face limited liability for downstream harms. Second, addressing systemic risk is made more difficult by a public goods problem: The consequences of a breach applies across organizations. No single actor has an incentive to internalize the costs of a systemic breach, and the result will be an inefficiently low level of investment by institutions to reduce these risks. Third, information technology product markets

are concentrated for various reasons: network effects, natural market power associated with high fixed costs and very low marginal costs in software, and regulatory or other barriers to entry. As a result, any cybersecurity flaw is therefore likely to create widespread impact. Finally, policy responses are most effective when they target the underlying market failures directly rather than focusing exclusively on compliance requirements.

Economics and Systemic Cyber Risks: The SolarWinds Case Study

We begin our analysis with a case study of a cybersecurity event at the SolarWinds company in 2020, which was a huge shock to the cybersecurity ecosystem.

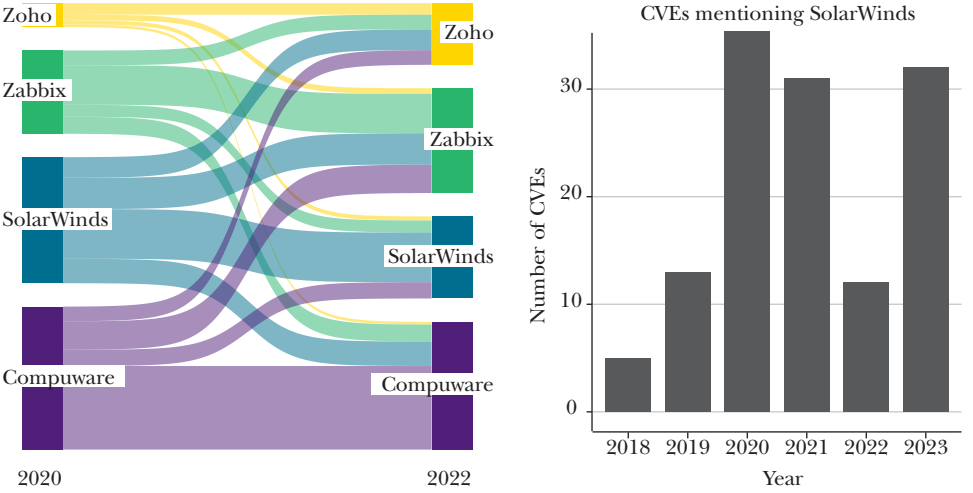
In 2020, the Orion platform of enterprise software supplier SolarWinds was hacked (Tidy 2020). In all, over 18,000 organizations in 36 countries downloaded the malicious code, including many governments and large enterprises with seemingly sophisticated cybersecurity programs in place—including the US Departments of Treasury and Commerce and a majority of Fortune 500 companies. This software, which, previously, few people had even heard of, was deployed by many of the world's largest corporations and governments to monitor their most critical information technology assets. Thus, the vulnerability granted unauthorized access to the most sensitive systems at thousands of high-value targets in an instant.

Attackers gained unauthorized access to the SolarWinds "build platform/system," used to compile and package the Orion software. In this case, the build system was the critical point of failure that enabled the attacker to inject malicious code into the compiled SolarWinds Orion software. When users installed the compromised software, the malicious code was executed on their systems, giving the attacker unauthorized access to their systems. The attacker was then able to steal sensitive data from their systems, such as credentials, intellectual property, and customer information.

Afterward, other significant vulnerabilities in the SolarWinds software and operational practices came to light. Perhaps most egregiously, it was found that they had used a default password of "solarwinds123" for configuring their systems, which the chief executive officer blamed on an intern (Fung and Sands 2021). However, there had been earlier warnings about cybersecurity at SolarWinds. In the year before the breach, SolarWinds was ranked by Intangic (a firm that estimates cyber risk, often for insurance purposes) as a high-risk company relative to its competitors, in part because SolarWinds had relatively high spending on mergers and acquisitions, but relatively low research and development spending (as reported by Nolan and Fixler 2021).

In spite of weak security practices, the SolarWinds Orion platform operated for many years as a leading provider in a niche area of corporate information technology software. How did this come to be? First, like many technology products, network effects tend to produce dominant products within market segments. There

Figure 1
SolarWinds Before and After the Hack



Source: Authors' calculations based on databases we discuss below.
Note: Sankey diagram (left) shows the migration of firms between vendors between 2020 and 2022. The vertical bars represent market shares. The bar plot (right) shows the total common vulnerabilities and exposures added to the National Institute of Standards and Technology database each year that mentioned SolarWinds in the description.

are benefits to scale in enterprise software, just as we have seen in operating systems (say, Microsoft Windows for desktop computing and the IOS and Android operating systems for mobile phones). Second, the enterprises that purchase software lack the expertise to ascertain the security quality of the products they are purchasing. This information asymmetry enabled SolarWinds to invest fewer resources into securing their products, as their customers could not judge whether they were more secure than their competitors. Third, the breach generated huge negative externalities in that thousands of SolarWinds customers were harmed by the breach much more than the vendor itself.

One might expect that the market would provide an incentive for SolarWinds to invest more in security, because customers would respond to a hack by switching to competitors. In fact, we see only limited movement away from SolarWinds after the attack. We identified three major competitors to SolarWinds Orion product using Gartner Peer Insights (Gartner 2026): Compuware Dynatrace, Zoho ManageEngine, and Zabbix. Because the attack was not put into place (via updates) until early 2020 and was not discovered for many months, it makes sense to consider the 2020 data on product installations to be the "before" period, while 2022 is the "after" period.

The market share of enterprises with each of these products in 2020 and 2022 is shown in Figure 1 (left). Within two years of the SolarWinds hack, several of its

competitors did experience some growth, while the installed base of SolarWinds Orion did decline somewhat. Nevertheless, as Figure 1 (left) shows, the market reaction was limited and the SolarWinds product was still used by a large number of enterprises several years after the attack.

A plausible rationale for the limited change is that in software, the switching costs to move from one enterprise software to another are usually high. SolarWinds Orion is typically deeply embedded in the information technology environment of an enterprise; changing to a competitor is a substantial undertaking that introduces new risks that the executives in charge of information systems may not want to take on.

Another possible explanation for why more firms did not switch could be that SolarWinds improved its security practices after the incident. The evidence does not support this explanation. As discussed above, the National Vulnerability Database posts known vulnerabilities in software products. Although SolarWinds patched the Orion system after the attack and released new versions, remarkably (and unfortunately), SolarWinds had nearly as many common vulnerabilities and exposures reported in 2023 as they did in 2020 and 2021, as shown on the right-hand side of Figure 1.

It is also instructive to consider what policy remedies were available and how they fared. First, firms like SolarWinds can disclaim any liability for defects in their software products, which naturally dulls any incentive to proactively seek out and fix defects. Nonetheless, the Securities and Exchange Commission (SEC) requires publicly traded companies to disclose cybersecurity practices and risks. In 2023, the SEC charged SolarWinds and its chief information security officer with fraud and internal control failures. The SEC claimed that in filings up until December 2020, SolarWinds downplayed its cybersecurity risks (and misled investors) by disclosing only generic and hypothetical risks, even though SolarWinds internal documents assessed that its product was not very secure, which was consistent with the external assessment of the Foundation for Defense of Democracies and Inganic (US Securities and Exchange Commission 2023). However, the SEC dropped the charges in 2025, which unfortunately sends a signal to firms that the SEC cybersecurity reporting rules are not very effective (US Securities and Exchange Commission 2025).

Given the absence of regulatory remedies, one might expect that the private market for cyber insurance could help to manage the risks. In fact, insurers also bore only a small fraction of the overall damages (about 1 percent by some estimates). Policies were written in such a way that the incident was either excluded from coverage or they did not trigger a claim.

This case study highlights the systemic risk created by market concentration; the incidence of the costs of the attack across software customers, vendors, and insurance; and the apparent failure of market incentives to improve security in the years after the consequences of the incident became apparent. The case further illustrates that market incentives do not seem to have mitigated the systemic cyber risk created by market concentration.

Measuring Systemic Cyber Risk

Perhaps surprisingly, there is no universally accepted definition for systemic cyber risk (Forscey, Bateman, Beecroft, and Woods 2022).We use the term "cyber vulnerability" to refer to the vulnerability of a single entity to cyber attacks. Cyber vulnerability can manifest through security flaws in software or underinvestment in defenses by organizations. We can then build on entity-level vulnerability to define systemic cyber risk, both *product-level systemic cyber risk* arising from concentration among technology suppliers that have significant cyber vulnerability and *firm-level systemic cyber risk* when dominant firms in a sector exhibit significant cyber vulnerability.

In this section, we seek to measure cyber vulnerability both at the product-level and firm-level. Notably, the Cybersecurity and Infrastructure Security Agency (2024), an agency within the US Department of Homeland Security, has called for measuring such sources of systemic cyber risk: "[S]imply put, what gets measured can be managed. CISA's goal to reduce systemic cyber risk is centered around finding concentrated sources of risk that, if mitigated, not only provide the organizations cost benefits for heightened risk management but also manage critical risks to the Nation's security and economic security."

Many firms and sectors within the global economy depend on a common set of critical technology products and software services, thereby consolidating risk across a small number of potential failure points. This convergence has undoubtedly enhanced efficiency, expanded capabilities, and increased flexibility for many organizations. However, the widespread reliance on a limited set of specialized products and services has simultaneously given rise to many single points of system vulnerability. Such concentration of risk amplifies the potential fallout from security breaches, especially since attackers have strong incentives to specifically attack such products and services.

In this section, we develop a methodology for measuring systemic cyber risk using available data. We do this by combining two types of datasets we have assembled: (1) installations of information technology/security products at the firm/site level and (2) vulnerability data at the product level. The dataset on information technology security dataset provides both detailed data on firms that sell cybersecurity products and on firms that adopt cybersecurity products. We have data from the Spiceworks Company Intelligence Database (CIDB) (Spiceworks Ziff Davis 2026) on worldwide firms spanning three years (2018, 2020, 2022). The data are on 11,000 installed information technology products at the entity/site/class/subclass/manufacturer/model level. Each of the observations includes detailed information about the entity/site including name of firm and site, location, and industry (by Standard Industrial Classification code). Each observation has a single product that is installed at that entity/site.

The takeaway for economics researchers is that it is possible to measure systemic cyber risk throughout the economy and to draw lessons that can be used to guide policy. In the following subsections, we show how this can be done at both the product level and the firm level.

Table 2
Market Shares of the Leading Suppliers in Routers: 2018, 2020, 2022

Year	Vendor market share (routers)								HHI
2018	Cisco	55%	Mikrotik	21%	SMC	12%	Dell	3%	3,667
2020	Cisco	51%	F5	8%	Mikrotik	7%	Asus	4%	2,830
2022	Cisco	51%	F5	13%	Juniper	6%	Aruba	4%	3,257

Source: Spiceworks Ziff Davis (2026).
Note: The Herfindahl–Hirschman Index (HHI), a standard measure of industry concentration, is defined as the sum of the squares of the market shares. According to Department of Justice guidelines (US Department of Justice 2024), an HHI index less than 1,000 usually means a very competitive market, the 1,000 to 1,800 range typically means a moderately concentrated market, and above 1,800 is usually considered to be concentrated.

Measuring Product-Level Systemic Cyber Risk: The Case of Cisco

Cisco is the market leader in routers, with more than 50 percent of this market. Routers form the backbone of the Internet; they are used to direct traffic across internet service providers, as well as within enterprises. Table 2 shows the market shares of the leading suppliers in routers in 2018, 2020, and 2022 in the United States, and shows the Herfindahl-Hirschman measure of industry concentration for each year.

We built a dataset of (1) common vulnerabilities and exposures per vendor in the router category and (2) the installed base of routers per vendor, using the Spiceworks Company Intelligence Database data. Including Cisco, the correlation between CVEs and installed base is 0.89, a *very* high positive correlation. Because Cisco has such a large installed base (and many router vulnerabilities), we also calculated the correlation with Cisco excluded, which remains quite significant at 0.46. These correlations suggest that the presence of systemic cyber risk in routers, because there is both high cyber risk (in terms of many vulnerabilities in the leading products) and high systemic risk (in terms of product concentration).

Attacks via Cisco routers have caused widespread systemic damage. The "Salt Typhoon" cyber event, widely attributed to China, has been described as the "worst telecom hack" in US history. Though the attack likely began in 2022, it was only discovered at the end of 2024. Cisco routers were specifically targeted and Salt Typhoon attackers gained access to telecoms through Cisco devices. According to US officials, the objective was to provide Chinese operatives with sustained access to US telecommunications networks by compromising devices such as routers and switches. A report by Cisco Talos—Cisco's threat intelligence unit—noted that in one instance, the Salt Typhoon attackers likely gained access to Cisco devices by exploiting a seven-year-old critical vulnerability in Cisco's IOS XE operating system (as reported by Kapko 2025a). Additionally, Recorded Future—a subsidiary of Mastercard that provides global threat assessments—wrote that five telecom networks were hit by Salt Typhoon via two other vulnerabilities in the Cisco IOS XE operating system at the end of 2024. They write that the attackers targeted

over 1,000 Cisco routers globally during late 2024 and early 2025, with a primary focus on telecommunications networks (Kapko 2025b; Recorded Future Insikt Group 2025).

Firm-Level Systemic Cyber Risk

We can also measure systemic cyber risk at the firm level. A promising start is the five categories of security controls from the National Institute of Standards and Technology Cybersecurity Framework (CSF): identify, protect, defend, respond, and recover (National Institute of Standards and Technology 2024). A brief description of the five categories of controls can be found at the NIST website (https://www.nist.gov/cyberframework/getting-started/online-learning/five-functions). We can assign security products by function to one of more of these five categories. We can then examine the number of categories covered by firms.

Of the 24 product series categorized by the Spiceworks Company Intelligence Database, three are specific to security: security devices, information technology security information, and network security monitoring. These three product series include a total of 507 security products. Data from the CIDB enable estimation of firm-level cybersecurity posture based on the number of security controls utilized by firms.

Using 2022 data, Table 3 illustrates the type of data that we can employ to measure cyber risk at the product and firm level by category. The first column is the twelve Spiceworks Company Intelligence Database security categories. The second column is our mapping of the CIDB security categories to the five National Institute of Standards and Technology Cybersecurity Framework functional categories. Thus, we mapped all 507 security products to the five NIST CSF categories. The third column of Table 3 is the number of products in each category. Column 4 is the percentage of firms that have installed a product in the category.

Table 3 also reports a count of common vulnerabilities and exposures and a list of known exploited vulnerabilities by product category from the US Cybersecurity and Infrastructure Security Agency. Column 5 is the total number of CVEs in products in the category, while column 6 is the total number of KEVs in products in the category.

Firm Coverage of NIST Cybersecurity Framework Categories

From this data, we can calculate the percentage of firm sites that cover each of the five National Institute of Standards and Technology Cybersecurity Framework categories. (Many firms have multiple sites.) Of the firms observed to employ at least one security control, just 39 percent of sites have controls covering two categories, while 21 percent cover three of the categories. Just 10 percent of the sites cover four of the categories, while only 2 percent of firms have controls covering all five categories. This is surprising, given that NIST recommends that all organizations adopt a baseline set of controls in all five categories. Even if we restrict attention to large firms with large sites with more than 500 employees, only 21 percent of the sites cover four of the NIST categories, while 6 percent have controls in all five

Table 3
Security Product Categorization and Installation Metrics

Category	NIST CSF category	Number of products	Percent of firms	CVEs	KEVs
IT asset management	Identify	8	3.7%	49	0
Advanced threat protection	Protect	18	11.6%	125	0
Data loss prevention	Protect	4	5.0%	14	0
Firewall software	Protect	35	26.9%	1,237	28
Identity access management software	Protect	63	45.8%	615	3
VPN	Protect	17	30.5%	858	15
Antivirus	Detect	138	17.6%	803	95
Email security	Detect	7	46.7%	72	0
Endpoint security	Detect	42	22.2%	1,467	24
SIEM	Respond	85	26.6%	711	2
Disaster recovery	Recover	27	11.6%	39	0
Other security	Multiple	33	23.4%	232	0

Source: Data source: Spiceworks Ziff Davis (2024). Table source: Hajizada, Adams, and Moore (2024).
Note: NIST CSF refers to the Cybersecurity Framework from the National Institute of Standards and Technology. SIEM refers to "security information and event management" products. CVEs are common vulnerabilities and exposures and KEVs are known exploited vulnerabilities.

categories. It is likely that some firms have greater coverage than reported here, but the overall picture remains troubling.

Following this discussion, we can now expand a bit on our earlier working definition of firm-level systemic cyber risk: Large firms that employ many employees and serve many markets are a source of systemic cyber risk if they do not have security products in at least three of the five of the National Institute of Standards and Technology Cybersecurity Framework categories.

Sector Coverage of the National Institute of Standards and Technology Cybersecurity Framework Categories

We can aggregate from the firm level to the sector level by using the North American Industry Classification System (NAICS). This system is used by US government statistical agencies to classify firms by sectors. There are 96 three-digit subsectors in the NAICS structure. This delineation is known as the NAICS3 level.

Table 4 presents the ten lowest-performing sectors in terms of the share of firms implementing controls in at least three of the Cybersecurity Framework categories at the NAICS3 level. In the table, the first column is the number of firms in the NAIC3 sector, while column 2 is the percent of firms in the sector that employ products in least three of the five National Institute of Standards and Technology categories. Column 3 is the average number of NIST categories covered by firms in that sector. Finally, column 4 is the name of the sector.

Table 4
Security Control Adoption Level for the Ten Worst-Performing Sectors

NAICS3	≥ 3 NIST (percent)	Average number NIST	Sector name
238	6.5%	1.32	Specialty trade contractors
624	7.5%	1.37	Social assistance
712	7.5%	1.36	Museums, historical sites, and similar institutions
321	9.2%	1.43	Wood product manufacturing
337	9.3%	1.44	Furniture and related product manufacturing
713	10%	1.40	Amusement, gambling, and recreation industries
484	11%	1.52	Truck transportation
332	11.1%	1.46	Fabricated metal product manufacturing
925	11.9%	1.51	Administration of housing programs, urban planning, and community development
611	12.0%	1.46	Educational services

Source: Hajizada, Adams, and Moore (2024)
Note: NAICS3 refers to the North American Standard Industry Classification at the three-digit level. NIST refers to the National Institute for Standards and Technology. Sectors are based on the Spiceworks Company Intelligence Database dataset for 2022. The label "≥ 3 NIST (percent)" is the percentage of firms in the sector employing products in at least three NIST categories. The heading "Average number NIST" is the average number of NIST categories covered by firms in that sector.

Overall, only 29 percent of firms in NAICS3 sectors implement controls in three or more of the five categories. Especially concerning are truck transportation and educational services. In these sectors, less than 12 percent of the firms had products in at least three of the five Cybersecurity Framework functional categories. While Table 4 shows that some other NAICS3 sectors have even lower levels of implementation, we focus on these two because of their importance: Transportation and manufacturing are included in the 16 critical infrastructure sectors, and education is extremely important as well.[1] In these two NAICS3 sectors, column 3 of the table shows that the average number of CSF controls employed is only roughly 1.5 out of the 5 CSF categories. These results show that we can identify sectors with high cyber risks. The next step is to check these sectors for firms with concentrated market share, which would indicate elevated systemic cyber risk.

A Lagging Industry Sector

We take a closer look at one industry sector that has especially lagged in cybersecurity investment: municipalities, which are critical infrastructure services. (The

[1] The sixteen "Critical Infrastructure Sectors" are defined in the National Infrastructure Protection Plan (US Department of Homeland Security 2013). They are as follows: chemical sector; commercial facilities sector; communications sector; critical manufacturing sector; dams sector; defense industrial base sector; emergency services sector; energy sector; financial services sector; food and agriculture sector; government services and facilities sector; health care and public health sector; information technology sector; nuclear reactors, materials, and waste sector; transportation systems sector; and water and wastewater sector.

Spiceworks Company Intelligence Database database includes municipalities.) Municipalities typically provide water and wastewater services.

Like many commercial enterprises, governments have increasingly employed digital technologies to supply services. While this undoubtedly brings benefits, there are costs as well, including increased exposure to cyber attacks. Whereas federal and state governments are reasonably well equipped to protect against cyber threats, cities and towns are not (Cybersecurity and Infrastructure Security Agency and Federal Bureau of Investigation and National Security Agency 2022).

Several recent reports suggest that ransomware attacks against municipal governments have increased significantly in recent years. A 2022 survey from the Multi-State Information Sharing and Analysis Center (MS-ISAC) found that ransomware incidents against municipalities had increased by 51 percent (Fox-Sowell 2024). MS-ISAC's 2025 report also found a significant increase in cyber attacks on critical municipal infrastructure in 2024 MS-ISAC (2025).

In July 2026, sophisticated hackers attacked many US municipal water systems in seven states. In this case, US and state officials believe that the attacks were by Iranian hackers. In several cases, hackers managed to change passwords and network configurations. This meant that the municipalities could not access the automated control systems that manage pumps, valves, and other infrastructure. Most of the more than 50,000 water systems in the United States serve less than 10,000 people. Many depend on decades-old infrastructure that were not engineered for a networked environment where ransomware and state-sponsored attacks pose serious problems (Volz and Londoño 2026; Easterly 2026). It is likely that the systemic cyber risk associated with the use of common infrastructure enabled so many simultaneous attacks.

Fully 80 percent of US municipalities reported having fewer than five employees dedicated to cybersecurity. In order to cope, many municipalities are switching to third-party providers including "managed service providers" to address growing information technology and cybersecurity needs. Several of these managed service providers serve many municipalities. While these services can be helpful for municipal cybersecurity, they also creates a common point of vulnerability across customers.

We examined the security technologies for all US government organizations gathered by the Spiceworks Company Intelligence Database in 2022. Table 5 cuts the data in two ways. First, it counts the number of the National Institute of Standards and Technology Cybersecurity Framework categories covered by each governmental unit. Second, it compares the overall results for local, state, and federal governments. We can see that municipalities badly lag state and federal governments. Of the municipalities in the CIDB database, 23 percent have at least one of the five categories, but it falls off quickly after that. Just 5 percent of municipalities have controls covering two categories, while 2 percent cover three of the categories. Just 1 percent of the municipalities cover four of the NIST categories, while only 0.2 percent of municipalities have controls covering all five categories. While some of the poor performance can be attributed to gaps in CIDB coverage, these findings underscore how far behind municipalities are in terms of cybersecurity technology adoption.

Table 5

Governmental Coverage of the Cybersecurity Framework Categories of the National Institute of Standards and Technology

	All governments		Municipality		State		Federal	
Group	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Identify	969	2%	236	0.3%	498	12%	235	12%
Protect	10,368	17%	7,080	13%	2,372	57%	916	48%
Detect	9,877	17%	7,095	13%	2,016	49%	766	40%
Respond	3,749	6%	1,432	3%	1,677	41%	640	34%
Recover	1,822	3%	622	1%	764	18%	436	23%
$\geq$ 1 NIST category	15,934	27%	12,236	23%	2,655	64%	1,043	55%
$\geq$ 2 NIST categories	5,557	9%	2,770	5%	2,043	49%	744	39%
$\geq$ 3 NIST categories	3,096	5%	951	2%	1,540	37%	605	32%
$\geq$ 4 NIST categories	1,535	3%	371	1%	763	18%	401	21%
5 NIST categories	663	1%	137	0.2%	326	8%	200	10%

Source: Spiceworks Ziff Davis (2026).
Note: Based on the Spiceworks Company Intelligence Database dataset for 2022.

Summary: Measuring Systemic Risk

Cyber risk can be measured, both at the product level and the firm level—and measurement is a first step in managing such risk. In particular, systemic cyber risk arises in settings with concentration. The literature on innovation and market structure in economics identifies a trade-off: Concentration enables scale economies and allows firms to profit from their innovation, but oligopolistic market structures can lead to low innovation due to lack of competitive pressure. The empirical evidence presented here is not sufficient to demonstrate a causal link, but it is consistent with a belief that, when it comes to cyber risk, concentration is not creating sufficient investment incentives to alleviate systemic risk or overcome free riding, policy failure, oligopoly behavior, and the lack of customer responsiveness to poor quality.

Policy Options for Mitigating Systemic Cyber Risk

Having demonstrated the widespread presence of systemic cyber risk and its economic drivers, we now turn to discuss policy interventions that could counteract the risks. We begin with traditional regulatory approaches and illustrate their limitations in this context. We then turn to nontraditional mechanisms that can help promote cybersecurity public goods and address information issues driving systemic cyber risk. Finally, we consider the potential and pitfalls of insurers in mitigating systemic cyber risks.

Throughout the discussion, we point to the economic forces that are critical in guiding policy. We also present several policy recommendations that we believe will improve cybersecurity, streamline regulation, and increase competitiveness in

government procurement, which will reduce market concentration and help mitigate systemic cyber risk and reduce its costs.

Promoting Secure Practices

When faced with problematic behaviors, such as widespread insecurity of software or technology products, policymakers traditionally respond in one of two ways: by enacting regulations in advance intended to limit such behavior or by imposing liability after undesirable events have occurred. Given US rules about tort liability, a policy of unlimited or joint liability for software makers could easily make developing new software, even software intended to fix cyber problems, into an investment that no firm would wish to undertake. It seems especially difficult to impose liability for a product that seemed safe when issued and tested, but that later a hacker found a point of entry.

Cybersecurity safety regulation has fared a little better. Policymakers have proposed instituting baseline security requirements for operators in critical infrastructure industries. This approach has recently gained traction in the European Union via the Cyber Resilience Act, which broadly requires that manufacturers follow cybersecurity rules in the planning, design, development, and maintenance of their products—sometimes including assessment by a third party. In July 2026, the European Commission published practical guidance for compliance with these rules.

One danger with EU-style detailed regulation—especially in the context of the fast-moving array of technologies potentially affected by cybersecurity concerns—is that time, energy, and resources can be wasted on negotiating such rule and accompanying compliance procedures without understanding whether there are corresponding benefits. Such a focus crowds out sensible "best practice policies"and creates fixed costs that act as a barrier to enter for small- and medium-size firms. For example, Weishäupl et al. (2018) find that cybersecurity investment decisions are frequently driven by compliance obligations, with learning about investment effectiveness occurring in an improvised and largely reactive manner. Related evidence suggests that firms may also rely on peer behavior when making security decisions: A study of Finnish information security managers documents herding behavior in cybersecurity investment, with firms following a "let's follow others" compliance strategy when selecting security controls (Shao, Siponen, and Liu 2020).

From Compliance Obligations to Best Practices: A Case Study

Until very recently, the US government was using a "compliance obligations" approach to cybersecurity for defense contractors: All US defense contractors were supposed to fully implement the more than 100 security requirements included in the National Institute of Standards and Technology Special Publication 800-171 by November 2026, often referred to as the Cybersecurity Maturity Model Certifications. These rules were to apply whenever a system owned or operated by or for a contractor processes, stores, or transmits "controlled unclassified information"—a regulatory category created by Executive Order 13556, issued in 2010, that refers

to the case of nonclassified information where the government still wishes to safeguard and limit or prevent dissemination.[2]

No empirical justification was given for why these more than 100 requirements were chosen. Implementation was required for all firms. This policy imposed large costs, but benefits were not obvious, and innovation incentives were near invisible. Unsurprisingly, adoption had been minimal, even for large firms that could absorb the costs.

During a "Promoting Competition in AI" conference at Stanford University in 2024, former US National Security Advisor and Secretary of State Condoleeza Rice was asked, "how you think the US government is doing today with respect to being able to leverage the commercial sector to acquire the technologies that are clearly going to be needed for national security purposes . . .". She replied as follows: "I think we're doing very poorly. Very poorly. And it's largely because the procurement and acquisition processes in the defense department are baroque to say the least."[3] She went on to note that it is virtually impossible for small companies with great ideas or products to"penetrate the Pentagon."

In this way, cybersecurity issues create barriers to entry and affect competition, especially in the defense procurement sector, but also across other sectors. This is because a number of departments, including NASA, the Department of Energy, and the General Services Administration (which acts as a centralized buyer for the federal government) all had various regulatory changes in process to roll out mandatory implementation of the NIST 800-171 standard. Indeed, this standard was to apply for all US federal government contractors, because the standard underlies the protection of US "controlled unclassified information."

The overall result is that, rather than leveraging US defense contracting to take advantage of the ecosystem of innovative small- and medium-sized technologies to improve cybersecurity, these firms instead would have been shut out, and many potential important cybersecurity innovations that could help to reduce cyber risk would never have come into the market. This important economic cost is typically overlooked.

It is important to emphasize that NIST 800-171 was not a reasonable step forward in cybersecurity for any firm, including large ones. Although large firms could afford the cost of implementation, the low adoption rate shows that they did not believe that the cybersecurity benefits exceeded costs.

Fortunately, the US government recently pivoted to a best practices approach. In July 2026, the US Department of Defense suspended implementation of these more than 100 security requirements. Instead, it will propose new regulations that lower barriers for small and medium firms, and replace bureaucratic compliance

[2]For additional detail, see the Defense Federal Acquisition Regulation Supplement 252.204-7012 Safeguarding Covered Defense Information and Cyber Incident Reporting. It is available at https://www.acq.osd.mil/dpap/dars/dfars/html/current/252204.htm#252.204-7012.

[3]For a description of the conferences, see the Stanford website at https://siepr.stanford.edu/events/promoting-competition-ai. The quotation from Rice is taken from a transcript of the conference.

with "scalable, resilient cybersecurity measures." This is a best practices approach (for discussion, see https://dodcio.defense.gov/CMMC/).

One way to implement a "best practices" approach to scalable, resilient cybersecurity measures is to use explicit cost-benefit methodology. Barnir, Gandal, Moore, and Scott (2025) showed that the number of requirements included in National Institute of Standards and Technology Special Publication 800-171 could be reduced by more than 50 percent without sacrificing much value, where value is the difference between benefits and costs. This "best practice" methodology facilitates opportunities for streamlined enforcement, focusing effort on the most important risks—particularly on those that affect systemic cyber risks. Such an approach will benefit all innovative firms. It will also likely reduce market concentration by enabling entry and competition.

This is an example of how economic analysis matters, especially for innovation incentives. Publication 800-171 implicitly assumed that technology evolves exogenously and that regulation does not have any tools at its disposal to shape either the rate or direction of technological change.

Imperfect and Incomplete Information

Imperfect and incomplete information generate suboptimal and ineffective cybersecurity. We now discuss four types of information issues that cause problems.

First, there is incomplete information on losses experienced by third parties. An organization knows when it is under cyber attack, but they do not always know when their peers are under attack. Even when an organization knows an attack has occurred elsewhere, the magnitude of the resulting loss is usually unknown. As a result, organizations do not know the cost associated with cyber attacks until they are directly experienced, which in turn makes it harder for firms to justify proactive investment in cyber defenses before attacks happen. Such incomplete information can contribute to negative externalities in that firms underinvest in cybersecurity defenses, which, in aggregate, can create systemic firm-level cyber risk.

Second, there is imperfect information on the effectiveness of security controls. Organizations struggle to determine which technologies are more effective at countering threats. Controls are selected based on measurable factors including cost, ease of use, and integration into existing infrastructure. Determining which products do a better job at stopping future attacks is very hard to measure, so the choice is not made on this basis.

Third, there is imperfect information on the frequency of attacks and the relative likelihood of different threats being realized. This makes it hard for organizations to use more sophisticated risk management approaches that can take attack likelihood fully into account.

Fourth, there is imperfect information on the attackers themselves. It is often unknown from where attacks originate, what the attack's motivation is, and what the attacker's capabilities are.

What should governments do to address imperfect and incomplete information problems? The most promising solution is to create missing information markets

and take steps to help shape these markets. Such information sharing markets could include threat intelligence, detailed loss variation and compromise frequency reports, evaluations of security measures, and consistent incident reporting. Such information would typically constitute a public good; the details of how such goods could be disseminated will be discussed in the next section.

However, we need to be careful that implementation does not diminish incentives to conduct research and development that seeks to find new security threats. If we required sharing of all such information, the incentives to conduct this form of research and development would be reduced, because proprietary information is a source of competitive advantage. Nevertheless, certain information could be in the public domain.

Information sharing may also yield a procompetitive effect. Hajizada et al. (2026) found that cyber attacks are also associated with changes in vendor portfolios. Compared to nonattacked firms, attacked firms were more likely to shift to vendors with larger market presence and broader product scope. While these large vendors' products may be higher quality, it is also possible that the switch to them is due to incomplete and imperfect information about vendor product quality.

Promoting Cybersecurity Public Goods via Data Sharing

Cybersecurity has a public good aspect. Many of the steps to promote cybersecurity—once they are discovered and produced—can be shared broadly at a near-zero marginal cost. The problem arises when entities can benefit from this cyber-risk prevention without contributing, leading to a market failure where security is underprovided at a systemic level. Firms may underinvest in cybersecurity, expecting others (competitors, government, or even information technology service providers) to cover the costs of securing the overall infrastructure, especially critical infrastructure operators. A mixture of institutional innovation and public rules can help to limit the extent of underinvestment in these public goods.

As one example, consider the role of cybersecurity data on vulnerabilities, attack strategies, and defenses. Once made public, such data can be shared and used by many without diminishing its availability to others (that is, it is "nonrivalrous"). Just as public institutions provide detailed weather reporting for farmers (and other industries), public institutions can provide detailed data on how, for example, cyber-criminals are profiting from their incursions, and ensure regulation is aligned with reducing the potential profits from activities like fraud, identity theft, and ransomware. Data sharing across companies about fraud, suspicious accounts, and other yellow or red flags would be extremely beneficial. Cyber insurance companies could play an important role in information sharing here. (We discuss insurance below.)

While there is broad agreement that data sharing on cyber risk issues would be beneficial, organizations will be concerned about not making public data about customer lists or other business secrets. In response to these concerns, data sharing could be limited so that the sharing only happens between firms, as already happens with sector-specific Information Security and Analysis Centers. Another

option is to require disclosure to government regulators but not the public, as has been adopted in EU regulations. Some proposals suggest that mandated disclosure of cybersecurity incidents might be paired with some shield against liability for the damage caused. The challenge here is to build a regulatory framework for data sharing; for example, it would provide specific guidance about how third party companies can combine and share data across firms for cybersecurity prevention purposes. Governments could make a substantial contribution in defining rules for such a framework.

Recent innovations in artificial intelligence have also made it cheaper for good actors to create software tools that identify and fix vulnerabilities or test updates before releases. A variety of digital public goods could be procured by governments and provided to industry actors. Novel procurement approaches for this software could be used, including prizes, contests, or "paying for outcomes" whereby vendors could be rewarded based on outcomes such as a low number of economically damaging cybersecurity incidents.

A related approach to addressing the public good aspect of cybersecurity might be carried out by "bug bounty" platforms that utilize crowdsourcing to find potential software vulnerabilities; for an examination of how these platforms work, Zrahia, Gandal, Markovich, and Riordan (2024) offer a useful starting point. Bug bounty platforms are two-sided markets that host bug bounty programs and researchers and enable security researchers (sometimes called "ethical hackers") to be rewarded for finding software vulnerabilities. The programs are structured as tournaments where companies pay monetary rewards for unique vulnerabilities found. The two most prominent bug bounty platforms are Hacker1 and Bugcrowd. The popularity of bug bounty platforms is increasing, and some governmental agencies have begun to use them. For example, in 2021, the Cybersecurity and Infrastructure Security Agency launched a federal civilian enterprise-wide "crowdsourced" vulnerability disclosure policy platform (CISA 2021). Nevertheless, these platforms have so far remained relatively small. While public disclosure of vulnerabilities is the default recommended policy for public programs at Bugcrowd and Hacker1, it is not required. Systematic cyber risk could be lowered by ensuring that nearly all vulnerabilities discovered on these platforms are (at least eventually) publicly disclosed. As above, advances in artificial intelligence may fundamentally alter the behavior of such markets, as the cost of discovering new vulnerabilities is lowered.

A Private Insurance Approach to Systemic Cyber Risk

The private cybersecurity insurance industry offers another set of institutions that might reduce systemic cyber risk. Total US premiums for cybersecurity risk were $9 billion in 2024; premiums in the rest of the world were another $6 billion (National Association of Insurance Commissioners 2025). However, despite the likelihood that cyber risk is rising, the total paid in cybersecurity insurance premiums has been largely flat in the last few years. One likely reason is that insurers quite rationally want to avoid systemic risk because it can lead to correlated losses. They

have largely succeeded in writing cybersecurity insurance policies to avoid covering systemic risks, which of course makes such insurance coverage less valuable to purchasers.

Nonetheless, insurance has the potential to play a critical role in measuring and managing systemic cyber risk, because insurance firms can gain data and expertise from evaluating and responding to cyber events across firms. Moreover, insurance firms have incentives to engage in research and development of knowledge, processes, and software that can help to assess cyber risk.

The key to getting social benefits from cybersecurity insurance is to ensure that it drives efficiency through scale economies (and the improved capabilities they bring) coupled with better aligned incentives, where insurance firms have substantial profit and reputation at stake in identifying and remediating problems. For example, insurance firms can use analysis from firms like Intangic that rate cybersecurity posture, and they can assess vulnerability to ransomware attacks by measuring the time it would take for a firm to recover compromised databases from backup (where longer times or uncertainty about capabilities mean firms are more vulnerable to ransomware demands). Firms with better demonstrated recovery capabilities are lower risk for ransomware payouts and, thus, can be offered lower premiums. Insurance companies can use their experience to design more effective cybersecurity audits, and they can share learnings from prior incidents (such as unexpected challenges in recovering from backup) with insured firms.

The challenge is to propose regulatory or policy approaches that will further approaches that increase the incentives for cyber insurance providers to contribute more directly to improving the risk profile of the firms they insure, either through incentives or the provision or curation of specialized services that are efficient and effective at reducing risk. Along these lines, a White House (2024) report on "The National Cybersecurity Strategy Implementation Plan" acknowledges the imperative (in strategic objective 3.6) to deepen our comprehension of how potential catastrophic cyber incidents might impact the cyber insurance marketplace (White House 2024).

Any proposed measures to bolster private cyber risk insurance must take into account the effects on the reinsurance market as well. Systemic cyber attacks can generate losses that no single insurer can absorb alone. Reinsurance makes it possible for insurers to underwrite large and complex cyber risks; without it, the insurance industry's ability to respond to catastrophic cybersecurity events would be seriously compromised. Because of the prevalence of systemic cyber risk, more than 50 percent of global cybersecurity insurance premiums are passed along to a reinsurance company; for property and casualty premiums, the equivalent figure would be about 20 percent. Additionally, the cyber reinsurance market is highly concentrated: The five largest players underwrite roughly 62 percent of gross written premiums, while the top ten collectively control 87 percent of the market. Should any major reinsurer curtail or exit its cyber capacity, the consequences for primary insurers—and for the market as a whole—could be severe (National Association of Insurance Commissioners 2025; Willard 2024).

Conclusion

Systemic cybersecurity risks are large and likely rising, while countermeasures against those risks seem inadequate. From an economic perspective, the underlying dynamics seem clear enough: Cybersecurity efforts have a substantial public good component, and individual firms (whether firms at risk of hacking or firms providing insurance against hacking) lack sufficient incentive to take the necessary steps. Instead, many of them free ride on the efforts of others.

In our own work in this area, we have been surprised that there is not more mapping of the specific types of market failures associated with systemic cyber risk to policies that would realign incentives. It is perhaps unsurprising that noneconomists have not focused on the incentive issues; in the case of economists, perhaps this shortcoming arises because the topic may appear too tactical for publishable economics research.

Our policy recommendations to reduce systemic cyber risk can be summarized as follows. The extent of concentration, competition, and interconnectedness affect the level of systemic cybersecurity risk and also the incentives of firms to address that risk. Incentives are shaped by market forces, institutional design, and other policies. We have pointed to areas where incentive alignment can be improved via policies, regulations, and private institutions such as insurance, but this essay is only a stepping stone toward the further research and concrete actions that are needed to address systemic cyber risk.

■ *Gandal and Moore gratefully acknowledge support from the US National Science Foundation (award 2147505 and award 2452738) and the US Israel Binational Science Foundation (award 2016622 and award 2021711). We are especially grateful to Timothy Taylor (JEP Managing Editor) for help in rewriting the paper as well as comments and suggestions that substantially improved the paper. We are also grateful to Heidi Williams (JEP editor), and Jonathan Parker and Jeffrey Kling (JEP coeditors) for important comments and suggestions that significantly improved the paper. We also thank Andrew Morin, Michael Riordan, Kostadis Roussos, and Vincent Scott for insightful discussions and suggestions.*

References

Barnir, Noa, Neil Gandal, Tyler Moore, and Vincent Scott. 2025. "A Cost-Benefit Approach to Optimizing Security Control Adoption." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4691404.

Check Point Research. 2026. *Cyber Security Report 2026*. Check Point Software Technologies Ltd.

CISA. 2021. "CISA Announces New Vulnerability Disclosure Policy (VDP) Platform." https://www.meritalk.com/articles/cisa-launches-new-vulnerability-disclosure-platform/.

Cybersecurity and Infrastructure Security Agency. 2024. "Systemic Cyber Risk Reduction." https://www.cisa.gov/resources-tools/programs/systemic-cyber-risk-reduction.

Cybersecurity and Infrasructure Security Agency. 2026. "Known Exploited Vulnerabilities Catalog." https://www.cisa.gov/known-exploited-vulnerabilities-catalog.

Cybersecurity and Infrastructure Security Agency and Federal Bureau of Investigation and National Security Agency. 2022. *Russian State-Sponsored and Criminal Cyber Threats to Critical Infrastructure*. Federal Bureau of Investigation.

Easterly, Jen. 2026. "Small Towns Shouldn't Have to Defend America's Water Supply from Iran." *New York Times*, August 2. https://www.nytimes.com/2026/08/02/opinion/iran-water-hack-trump.html.

Forscey, David, Jon Bateman, Nick Beecroft, and Beau Woods. 2022. "Systemic Cyber Risk: A Primer." Carnegie Endowment for International Peace, March 7. https://carnegieendowment.org/2022/03/07/systemic-cyber-risk-primer-pub-86531.

Fox-Sowell, Sophia. 2024. "Cyberattacks on State and Local Governments Rose in 2023, Says CIS Report." *StateScoop*, January 30. https://statescoop.com/ransomware-malware-cyberattacks-cis-report-2024/.

Fung, Brian, and Geneva Sands. 2021. "Former Solarwinds CEO Blames Intern for 'solarwinds123' Password Leak." *CNN*, February 26. https://www.cnn.com/2021/02/26/politics/solarwinds123-password-intern.

Gartner. 2026. "Browse 950+ Enterprise Software Categories." Gartner Peer Insights. https://www.gartner.com/reviews/markets.

Hajizada, Abulfaz, Dana Etgar Itzhaki, Noa Barnir, Tyler Moore, and Neil Gandal. 2026. "The Impact of Cyber Attacks on Security Portfolios: Firm-Level Panel Data." *CEPR Press* DP21219.

Hajizada, Abulfaz, Matthew Adams, and Tyler Moore. 2024. "Construction and Analysis of a Large-Scale Firm-Level Cybersecurity Posture Dataset." In *30th Americas Conference on Information Systems: Elevating Life through Digital Social Entrepreneurship*, AMCIS 2024, edited by Michelle Carter, Kelly J. Fadel, Thomas O. Meservy, Deborah J. Armstrong, Amit Deokar, Matthew L. Jensen. Association for Information Systems.

IBM. 2026. *Cost of a Data Breach 2026*. IBM.

Kapko, Matt. 2025a. "Salt Typhoon Gained Initial Access to Telecoms through Cisco Devices." *CyberScoop*, February 20. https://cyberscoop.com/cisco-talos-salt-typhoon-initial-access/.

Kapko, Matt. 2025b. "Salt Typhoon Remains Active, Hits More Telecom Networks via Cisco Routers." *CyberScoop*, February 13. https://cyberscoop.com/salt-typhoon-china-ongoing-telecom-attack-spree/.

Kunreuther, Howard, and Geoffrey Heal. 2003. "Interdependent Security." *Journal of Risk and Uncertainty* 26 (2): 231–49.

Mousqueton, Julien. 2026. "Ransomware.live." https://ransomware.live.

MS-ISAC. 2025. *Strengthening Critical Infrastructure: State, Local, Tribal, and Territorial Progress and Priorities Volume 1: Brief Report*. Multi-state Information Sharing and Analysis Center.

National Association of Insurance Commissioners. 2025. *Report on the Cybersecurity Insurance Market*. National Association of Insurance Commissioners.

National Institute of Standards and Technology. 2024. *The NIST Cybersecurity Framework (CSF) 2.0*. National Institute of Standards and Technology, US Department of Commerce.

National Institute of Standards and Technology. 2026. *National Vulnerability Database*. US Department of Commerce. https://nvd.nist.gov/ (accessed August 30, 2026).

Nolan, Chris, and Annie Fixler. 2021. "The Economic Costs of Cyber Risk." Foundation for Defense of Democracies, June 28. https://www.fdd.org/analysis/2021/06/28/the-economic-costs-of-cyber-risk/.

Recorded Future Insikt Group. 2025. *RedMike (Salt Typhoon) Exploits Vulnerable Cisco Devices of Global Telecommunications Providers*. Recorded Future.

Shao, Xiuyan, Mikko Siponen, and Fufan Liu. 2020. "Shall We Follow? Impact of Reputation Concern on Information Security Managers' Investment Decisions." *Computers and Security* 97: 101961.

Spiceworks Ziff Davis. 2026. "Data Intelligence." https://swzd.com/products/data/.

Tidy, Joe. 2020. "SolarWinds: Why the Sunburst Hack Is So Serious." *BBC News*, December 16. https://www.bbc.com/news/technology-55321643.

US Department of Homeland Security. 2013. *National Infrastructure Protection Plan*. US Department of Homeland Security.

US Department of Justice. 2024. "Herfindahl-Hirschman Index." January 17. https://www.justice.gov/atr/herfindahl-hirschman-index.

US Securities and Exchange Commission. 2023. "SEC Charges SolarWinds and Chief Information Security Officer with Fraud, Internal Control Failures." Press release, October 30. https://www.sec.gov/newsroom/press-releases/2023-227.

US Securities and Exchange Commission. 2025. *SEC v. SolarWinds Corp. and Timothy G. Brown*, No. 1:23-cv-09518-PAE (S.D.N.Y. filed Oct. 30, 2023). Litigation Release No. 26423.

Volz, Dustin, and Ernesto Londoño. 2026. "Scope of Hacks on U.S. Water Supply Widens as Evidence Points to Iran." *New York Times*, August 1. https://www.nytimes.com/2026/08/01/us/politics/iran-cyberattack-water-systems.html.

Weishäupl, Eva, Emrah Yasasin, and Guido Schryen. 2018. "Information Security Investments: An Exploratory Multiple Case Study on Decision-Making, Evaluation and Learning." *Computers and Security* 77: 807–23.

White House. 2024. *National Cybersecurity Strategy Implementation Plan, Version* 2. White House.

Willard, Jack. 2024. Ceded Reinsurance, an Ever Smaller Proportion of Total Premium over the Last Five Years: Howden Re." *Reinsurance News*, August 22. https://www.reinsurancene.ws/ceded-reinsurance-an-ever-smaller-proportion-of-total-premium-over-the-last-five-years-howden-re/.

World Economic Forum. 2026. *Global Cybersecurity Outlook 2026*. World Economic Forum.

Zrahia, Aviram, Neil Gandal, Sarit Markovich, and Michael Riordan. 2024. "The Simple Economics of an External Shock to a Bug Bounty Platform." *Journal of Cybersecurity* 10 (1): tyae006.

Journal of Economic Perspectives—Volume 40, Number 4—Fall 2026—Pages 137–158

In Praise of the Bayesian Bootstrap

Isaiah Andrews, Sarah Moon, and Jesse M. Shapiro

A central task in empirical science is to quantify uncertainty about our conclusions. For instance, how might our answer be different if we ran our survey, or experiment, again on a fresh sample? By and large, the default practice among economists is to quantify uncertainty by reporting a standard error or confidence interval built from an analytic formula.

Improvements in computing power have enabled other methods for uncertainty quantification, notably the bootstrap. In this article, we explain what the bootstrap is and how it works. We emphasize a particular kind of bootstrap, called the Bayesian bootstrap. The Bayesian bootstrap is just as convenient as the usual bootstrap but has many practical and conceptual advantages. We argue that the Bayesian bootstrap should be widely used within economics as a go-to method for quantifying uncertainty. We begin with two motivating examples that help to set the stage.

■ *Isaiah Andrews is the Charles E. and Susan T. Harris Professor of Economics, Massachusetts Institute of Technology, Sarah Moon is a PhD candidate in economics, Massachusetts Institute of Technology, and Jesse M. Shapiro is the George Gund Professor of Economics and Business Administration, Harvard University, all in Cambridge, Massachusetts. Andrews and Shapiro are also Research Associates at the National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are iandrews@mit.edu, sarahmn@mit.edu, and jesse_shapiro@fas.harvard.edu. Authors are listed in alphabetical order.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20261514.

Example A: Why Bootstrap?

Suppose we've run a randomized experiment in which we've randomly called a group of eligible citizens (the treatment group) by phone to encourage them to vote while another group of citizens received no call (the control group). Using voter registration files, we track which citizens actually voted. We estimate the difference in voting frequency between the treatment and control groups and we find it is positive—around a 5 percentage point difference. This is our point estimate of the treatment effect of calls on voting. We'd like to quantify the uncertainty in our point estimate.

The usual way to do that, taught in almost every introductory statistics course, is to assume that our estimate is more or less normally distributed and, more specifically, that if we repeatedly drew new samples from the same underlying population then, across these samples, the point estimate would more or less follow a normal distribution with mean equal to the true treatment effect and standard deviation equal to the standard error. Such an assumption can, in turn, be justified by a large body of theoretical work that shows that many kinds of statistics follow such normal distributions in large samples.

Panel A of Figure 1 shows what that theory implies in our case. Here, we take draws from a normal distribution whose mean is given by the point estimate of the effect of calls on voting and whose standard deviation is given by its standard error.

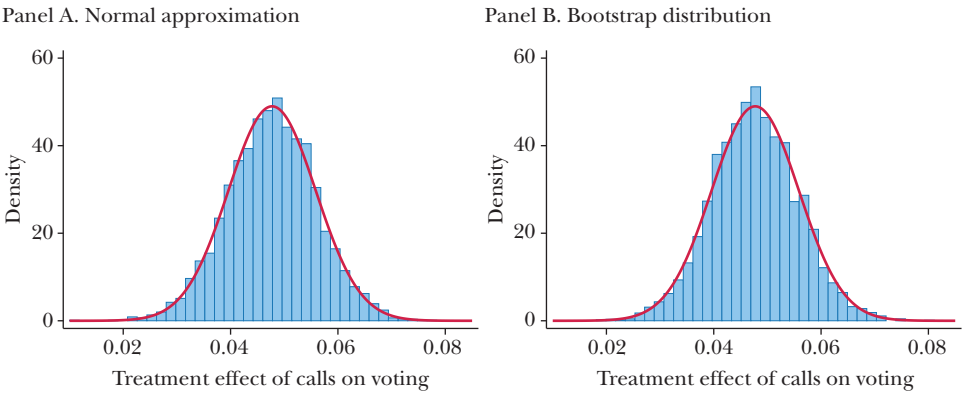
Is the normal distribution a good approximation of the uncertainty in the point estimate? As we detail below, one way to investigate is to run a bootstrap. To do this, we repeatedly create new, synthetic datasets by randomly sampling observations (citizens) from the original data with replacement. For each of these *bootstrap replicates*, we recompute the difference in voting frequency between the treatment and control group. Panel B of Figure 1 shows the resulting *bootstrap distribution*.

This bootstrap distribution looks pretty similar to the normal distribution with which we started! Indeed, they are very similar. In situations like this, the usual normal approximation performs pretty well. We can conclude with some confidence that the calls worked.

Does that make phone calls worthwhile for a political campaign? A different statistic of interest might be the cost per vote gained—that is, the cost in dollars of hiring campaign staffers, who will make those calls, divided by the number of votes gained by the calls. We know (roughly) the wage rate of campaign staffers and how many calls they can make per hour. We can divide the wage rate by the number of calls per hour, and divide that by the treatment effect of calls on votes, to get the implied cost per vote gained. Our point estimate is around $9 per vote gained.

How confident are we in this estimate? We can again consider the usual normal approximation, which we form by applying a procedure called the "delta method" to the normal distribution we used for the treatment effect. The delta method lets us calculate the standard error of a function—here, the wage per call divided by the treatment effect—of an already-estimated parameter—here, the treatment effect.

Figure 1
Quantifying Uncertainty for the Treatment Effect of Calls on Voting



Source: Calculations based on data from Nickerson, Friedrichs, and King (2006).
Note: Both plots show the density of the normal distribution with mean given by the point estimate of the treatment effect and standard deviation given by the standard error of the treatment effect. Panel A shows a histogram of 5,000 random draws from this normal distribution. Panel B shows a histogram of 5,000 bootstrap replicates of the estimated treatment effect. Data consist of 16,181 observations.

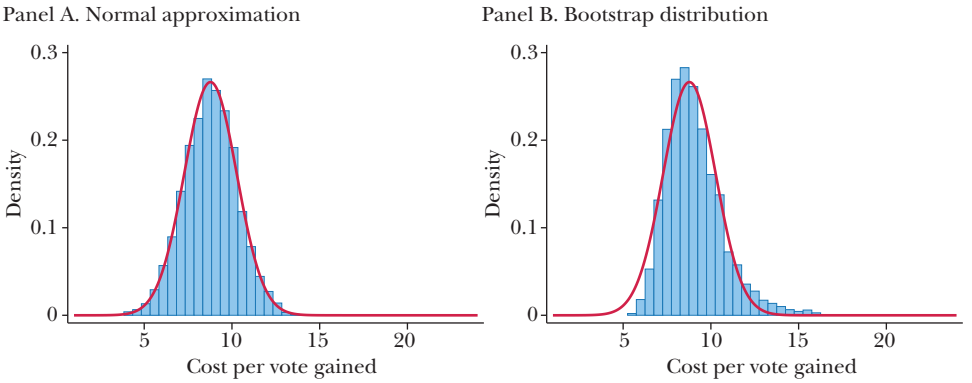
The delta method works by assuming that the function is approximately linear, which means that if the already-estimated parameter is approximately normal, then so (approximately) is the function's output, with the standard error of the output approximated by the product of the standard error of the already-estimated parameter and the derivative of the function. We can then take repeated draws from the resulting normal distribution with mean given by the point estimate of the cost per vote gained and standard deviation given by the standard error from the delta method. Panel A of Figure 2 shows the result.

The normal approximation says there is a reasonable amount of uncertainty about the cost per vote gained but suggests that the cost is below $14 or so. If other campaign methods (say, mailing leaflets) cost more than $14 per vote gained, then phone calls seem like a much better deal.

Is the normal approximation doing a good job of describing the uncertainty here? Let's investigate by running our bootstrap again, this time calculating the cost per vote for each bootstrap replicate. This is straightforward to do: For each replicate, we take the estimated treatment effect and then divide the wage rate per call by that treatment effect to get the estimated cost per vote. In panel B of Figure 2, we plot the resulting bootstrap distribution.

The result is visibly different! For example, whereas the normal approximation suggests that the cost per vote is less than $14, the bootstrap distribution suggests it may exceed $14. Even though the usual normal approximation works well for the difference in voting probabilities, it doesn't seem to be working as well for this very simple transformation of that difference.

Figure 2
Quantifying Uncertainty for the Cost per Vote Gained



Source: Calculations based on data from Nickerson, Friedrichs, and King (2006).
Note: Both plots show the density of the normal distribution with mean given by the point estimate of the cost per vote and standard deviation given by the delta-method standard error of the cost per vote. Panel A shows a histogram of 5,000 random draws from this normal distribution. Panel B shows a histogram of 5,000 bootstrap replicates of the estimated cost per vote. Data consist of 16,181 observations.

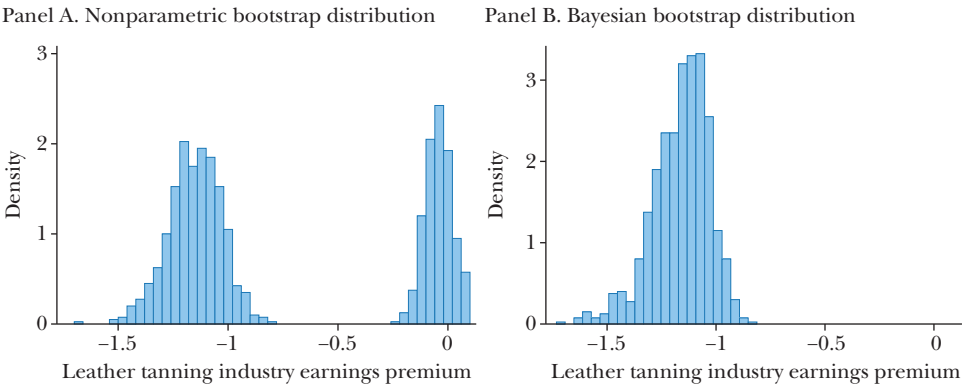
We are going to argue that, when the bootstrap distribution looks different from the normal distribution described by the point estimate and standard error, it usually means that something is wrong with the normal approximation. (Here, the delta method approximates the cost per vote as a linear function of the treatment effect, but the cost per vote is a ratio with the treatment effect in its denominator and is, therefore, very nonlinear for treatment effects near zero.) We are also going to argue that, when that happens, it is usually better to trust the bootstrap.

Example B: Which Bootstrap?

There are many ways to bootstrap. The most common one, which we've used above, involves repeatedly drawing random observations from the original dataset, as if they were tiles in a Scrabble bag (except that we put them back right after we draw them). This bootstrap, called the *nonparametric bootstrap*, works well in a range of situations, but it can sometimes get us into trouble.

Say, for example, that we are interested in the difference in earnings across industries. We know that different kinds of people—with different skills, education, and experience—work in different industries, so we cannot just compare different people working in different industries. Instead, let's look at how earnings change when the same person moves from one industry to another. We can do this by running a regression of the logarithm of earnings on indicators for industry, controlling for person fixed effects, time fixed effects, and other economic and

Figure 3
Quantifying Uncertainty for the Leather Tanning Industry Earnings Premium



Source: Calculations based on Integrated Public Use Microdata Series Current Population Survey longitudinal files from 1978 to 1980 (Flood et al. 2025).
Note: Panel A shows a histogram of 1,000 nonparametric bootstrap replicates of the leather tanning industry earnings premium, computed in StataNow 19.5. Panel B shows a histogram of 1,000 Bayesian bootstrap replicates of the leather tanning industry earnings premium, computed in StataNow 19.5. The leather tanning industry earnings premium is the difference in a worker's predicted log earnings from working in the leather tanning industry and working in their current industry, averaged across all workers. Data consist of 38,368 observations.

demographic characteristics. Krueger and Summers (1988) did something like this in a classic article on interindustry wage differentials.[1] Our estimating equation is

$$\ln(earnings_{it}) = \beta' Industry_{it} + \alpha_i + \lambda_t + \gamma' X_{it} + \varepsilon_{it},$$

where $Industry_{it}$ denotes a vector of industry indicators; α_i denotes person fixed effects; λ_t denotes time fixed effects; and X_{it} denotes a vector of controls for education, age, occupation, marriage, and their interaction with sex.

For any target industry, we can use the regression model to estimate the effect on log earnings of switching the average person from their actual industry into that target industry. When we do this for the leather tanning industry, we find that it is associated with very low earnings: On average, a worker in our sample would experience almost a 70 percent drop in earnings from moving from their actual industry to the leather tanning industry. Because we are wary of trusting normal approximations, we use a bootstrap to calculate the uncertainty in our estimate of the leather tanning earnings premium.

Panel A of Figure 3 plots a histogram of the estimated leather tanning earnings premium across the different bootstrap replicates. A strange pattern emerges with

[1] In this journal, Bonhomme, Lamadon, and Manresa (2026) discuss modern methods for a related class of problems.

two groups of points, which turns out to have a sinister cause. Recall that when we bootstrap, we draw observations (workers) at random so that, in any bootstrap replicate, some workers are included and some are excluded. Sometimes, it happens that we do not draw any workers who move between leather tanning and another industry. When that happens, we cannot estimate the premium to (not) working in leather tanning because of multicollinearity: For a worker who ever works in leather tanning, the leather tanning indicator is perfectly predicted by the worker's fixed effect, since neither varies with time. The software responds to the multicollinearity by automatically dropping the indicator for the leather tanning industry and assigning it a coefficient of zero, resulting in the cluster of points on the right.

There are many ways to get out ahead of problems like this, including being careful how we use our statistical software! It turns out, however, that there is a bootstrap that was invented over 40 years ago that avoids the problem altogether because it never drops any workers out of the sample. The distribution from that bootstrap looks like panel B of Figure 3.

This other bootstrap, called the *Bayesian bootstrap*, is just as convenient to run as the usual one. In addition to never dropping observations, it has many nice properties that make its output straightforward to interpret. For these reasons, we are going to argue that economists should be using it instead of the nonparametric bootstrap.

Some Background on the Bootstrap

Origins of the Bootstrap

The go-to way to quantify statistical uncertainty is to use an analytic formula that is intended to describe, at least approximately, the uncertainty in a statistic of interest. Think, for example, about the well-known formula for the standard error of a sample mean, which divides the sample standard deviation by the square root of the sample size. When such a formula is available, it is often very convenient and straightforward to use.

Sometimes, however, such analytic approximations are unknown or unavailable. Or sometimes, as in Example A, these analytic approximations do not perform all that well, because the assumptions on which they rely, such as approximate normality or linearity, can fail in real applications. These shortcomings led to the search for general-purpose alternatives that could approximate the uncertainty in a wide range of statistics under weaker assumptions.

For this purpose, Efron (1979) proposed what we have called the nonparametric bootstrap. This method involves repeatedly drawing from the sample with replacement and recalculating the statistic in which we are interested. The idea is to use the sample we have to guess what other possible samples might have looked like. Because the bootstrap uses the available sample to learn about other possible samples, it appears to do the impossible, much like "pulling ourselves up by our bootstraps."

In the next section, we summarize how the nonparametric bootstrap works. We discuss that it is an example of a more general type of bootstrap that involves randomly reweighting the data. We argue that this way of understanding the bootstrap helps clarify what it does, makes it easier to compute, and illustrates some downsides of the nonparametric bootstrap that other types of weighted bootstraps can fix.

The Bootstrap as a Reweighting of the Sample

The nonparametric bootstrap involves repeatedly drawing observations with replacement from the sample. Each set of draws makes up a bootstrap replicate. We then recompute the statistic of interest on each bootstrap replicate to obtain the bootstrap distribution, as in the examples above. (This description of the nonparametric bootstrap applies to the case of independent observations; we will discuss how to handle clustered observations later.)

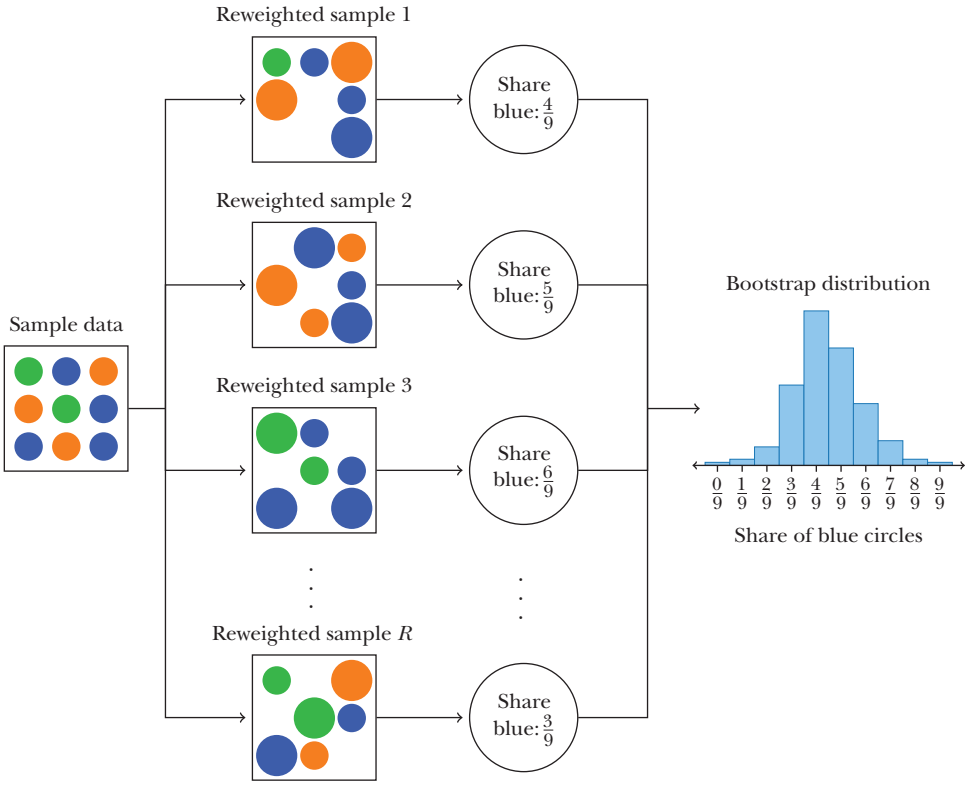
If you reflect on this process, drawing observations with replacement is equivalent to assigning integer-valued "frequency weights" to the observations in the original sample. An observation that was never drawn gets a weight of zero, an observation that was drawn twice gets a weight of two, and so on.

The nonparametric bootstrap is, therefore, an example of a more general class of bootstraps called *weighted bootstraps*. A weighted bootstrap proceeds by repeating the following two steps: First, randomly pick a set of weights; second, calculate the statistic of interest on the reweighted sample using those weights. What distinguishes the nonparametric bootstrap is simply the kind of weights it uses: the frequency weights. Formally, these frequency weights are a random draw from a multinomial distribution that assigns equal probability to each of the original observations. Just as the binomial distribution describes the number of heads (or tails) from repeated independent flips of a coin, the multinomial distribution describes the number of times a many-sided die lands on each of its sides in repeated independent rolls of the die.

Figure 4 visually illustrates the nonparametric bootstrap viewed as a weighted bootstrap on a sample of nine colored circles when the statistic of interest is the share of blue circles. We randomly draw R different sets of frequency weights to form R different reweighted samples. A box shows the weight assigned to each of the original nine circles in each reweighted sample. Here, a given circle gets a weight of zero (dropped from the box), one (included in the box in its original size), or two (included in the box at double its original size). On each of these reweighted samples, we calculate the share of blue circles, weighting each circle by its area. The collection of R-weighted statistics (shares) forms our estimate of the bootstrap distribution.

An advantage of this way of looking at the bootstrap is that it makes computation very convenient. If the statistic we're interested in (say, a regression coefficient) comes from a procedure (say, a regression command) that can accept weights for different observations (say, via a weights option in Stata), then all we need to do is to repeatedly call the procedure with each of the R sets of weights, as in the Algorithm 1 shown in Exhibit 1.

Figure 4
The Nonparametric Bootstrap as a Weighted Bootstrap



Source: Authors' rendering.
Note: The diagram illustrates how we can interpret the nonparametric bootstrap as a weighted bootstrap. From left to right, we resample with replacement R times to create R bootstrap replicates. We represent this resampling as reweighting: in a given replicate, a circle was chosen not at all (omitted, weight of 0), once (regular size, weight of 1), or twice (large, weight of 2). On each reweighted sample we calculate the share of blue circles by area. We then plot the distribution of the resulting collection of R shares.

Because the calculation for one set of random weights doesn't depend on the calculation for another set of random weights, we can calculate the statistic for the different bootstrap replicates in parallel. The parallel nature of the bootstrap, which is especially valuable in the modern era of parallel computing, is apparent in the shape of the arrows in Figure 4.

The reweighting perspective also highlights a key vulnerability of the nonparametric bootstrap. The pitfall we saw in Example B happens because when an observation gets a weight of 0 it effectively disappears, as happens in the reweighted samples in Figure 4. That can lead to all kinds of funny behaviors, one of which—being unable to estimate the model when certain observations are dropped in the bootstrap replicate—we saw in panel A of Figure 3.

Exhibit 1
The Coding Steps for a Weighted Bootstrap

Algorithm 1: Pseudocode for a Weighted Bootstrap
For each replicate 1…*R*, do:
- Draw pseudoweights for each observation according to some distribution
  - For example, for a nonparametric bootstrap, use multinomial
- Calculate weights by dividing the pseudoweights by their total across all observations
- Calculate replicate by applying the weights in your procedure
  - For example, for a least squares regression, do weighted least squares; for generalized method of moments, use a weighted moment function

Algorithm 2: Stata Pseudocode for a Single Replicate of the Nonparametric Bootstrap and Bayesian Bootstrap
Nonparametric bootstrap:
```
pseudoweight = colsum(rdiscrete(N, 1, J(N, 1, 1/N)) :== (1..N))'
reg vote treatment [fw = pseudoweight]
```

Bayesian bootstrap:
```
gen pseudoweight = rexponential(1)
reg vote treatment [weight = pseudoweight]
```

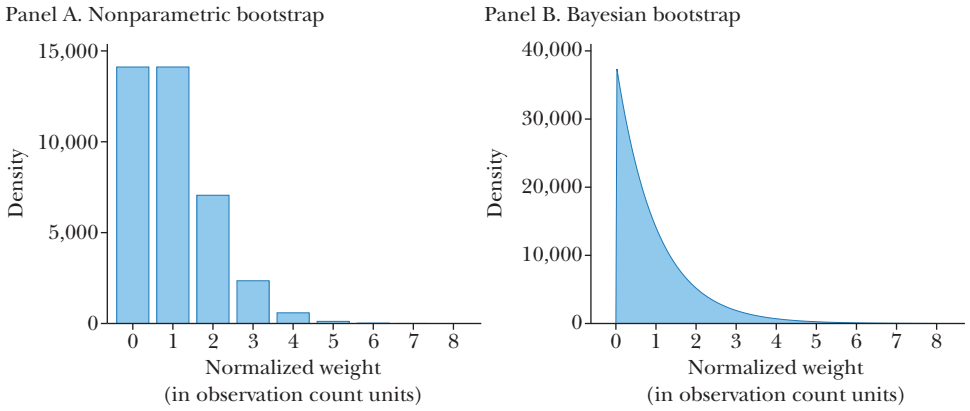
Algorithm 3: An Example of Stata Code for the Nonparametric Bootstrap and Bayesian Bootstrap
Nonparametric bootstrap:
```
bootstrap, reps(5000): reg vote treatment
```

Bayesian bootstrap:
```
bayesboot, reps(5000): reg vote treatment
```

Source: Authors' rendering.
Note: Algorithm 1 describes a generic procedure (with examples) for running a weighted bootstrap. Algorithm 2 illustrates the process of running a single replicate of the nonparametric and Bayesian bootstrap in Stata, where the statistic of interest is a regression coefficient, as in Example A. We first draw pseudoweights independently according to a multinomial distribution with uniform probabilities (nonparametric bootstrap) or a standard exponential distribution (Bayesian bootstrap). We then pass these as weights to the regression procedure. The step of normalizing the pseudoweights to sum to one is unnecessary because it is handled implicitly by the regression procedure. Algorithm 3 shows an example of Stata code that runs the nonparametric bootstrap and Bayesian bootstrap for the treatment effect of calls on voting in Example A. The `bootstrap,reps(5000)` command tells Stata to run the nonparametric bootstrap with 5,000 bootstrap replicates. To specify the statistic of interest, we add `reg vote treatment` after the colon, which tells Stata to regress `vote` on `treatment` in each bootstrap replicate and save each of the resulting coefficients. The command for the Bayesian bootstrap is the exact same, except that we use `bayesboot` to let Stata know to run the Bayesian bootstrap. (The `bayesboot` command was introduced in Stata version 19.)

These issues have been known in the statistics literature for a while—see, for example, Clyde and Lee (2001) and Xu et al. (2020). They pop up in economics too: Appendix B.1 of Angrist, Hull, Pathak, and Walters (2017), which is in turn referenced in Finkelstein, Gentzkow, and Williams (2021), mentions wanting

Figure 5
Comparing the Distribution of Weights Between Nonparametric and Bayesian Bootstraps



Source: Calculations based on Integrated Public Use Microdata Series Current Population Survey longitudinal files from 1978 to 1980 (Flood et al. 2025).
Note: The plots show the analytic densities of the bootstrap weights for the case of Example B, in which there are 38,368 observations. Panel A shows the density of the nonparametric bootstrap weights. Panel B shows the density of the Bayesian bootstrap weights. Weights are normalized to sum (or integrate) to one and are displayed in observation count units, so 1 denotes a normalized weight of 1/38,368. Densities are plotted in units of normalized weights.

to include "small lottery strata that would occasionally be dropped in a standard nonparametric bootstrap" as a reason not to use the nonparametric bootstrap. Footnote 24 of Lei and Ross (2025) says something similar.

Fortunately there is another type of weighted bootstrap that avoids this pitfall by using "smoother" weights. In the next subsection, we will introduce it.

The Bayesian Bootstrap

The Bayesian bootstrap was introduced by Rubin (1981) and was studied in economics by Chamberlain and Imbens (2003). For a more recent overview of the Bayesian bootstrap, Ghosal and van der Vaart (2017) offer a good starting point. The Bayesian bootstrap replaces the frequency weights of the nonparametric bootstrap with "smoother" weights. These smoother weights are proportional to draws from a standard exponential distribution, which means they are strictly greater than zero with probability one. Intuitively, these weights are a uniform random draw from the set of possible weights on the original sample: In the special case with two observations, the weight on each observation is distributed uniformly between zero and one.

Recall that, in the nonparametric bootstrap of Figure 4, the weights were either zero, one, or two. By contrast, under the Bayesian bootstrap, the weights are smoother, so that (intuitively) a circle can take on any area from small to large. Figure 5 illustrates this by plotting the distribution of the weights, in Example B, in

the nonparametric bootstrap (of panel A of Figure 3) and the Bayesian bootstrap (of panel B of Figure 3). The nonparametric bootstrap weights are discrete and often zero (panel A), whereas the Bayesian bootstrap weights are continuous and strictly greater than zero with probability one (panel B).

Other than how we calculate the weights, the Bayesian bootstrap and nonparametric bootstrap procedures are the same. Thus, many software implementations can treat them similarly. For example, Stata has similar implementations of both types of bootstrap. In Algorithm 3 of Exhibit 1, we show an example of Stata code that runs the nonparametric bootstrap and Bayesian bootstrap for the treatment effect of calls on voting in Example A. Under the hood, a single replicate of each bootstrap can be computed along the lines of Algorithm 2 shown in Exhibit 1.

From a practical standpoint, the main advantage of the Bayesian bootstrap is that because the weights are always positive, no observations are ever dropped. This means we can avoid some of the tricky situations we talked about in the last section and illustrated in Example B.

From a conceptual standpoint, the main advantage of the Bayesian bootstrap is that, in addition to inheriting many of the desirable properties of the nonparametric bootstrap, the Bayesian bootstrap has a Bayesian interpretation—hence the name. In the next subsection, we explain what that means and why we think it is valuable.

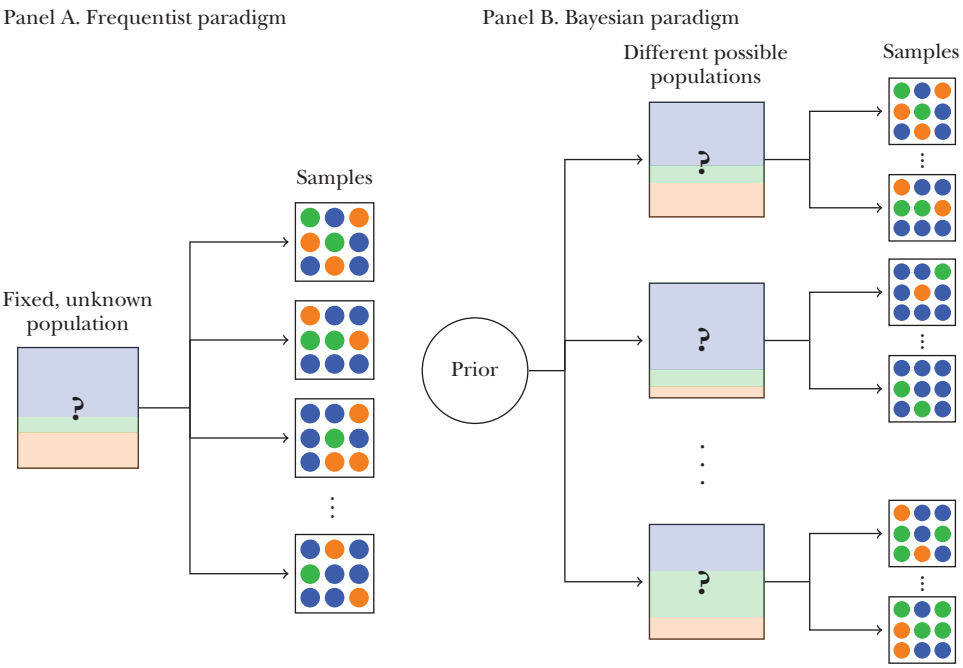
Frequentist and Bayesian Interpretations of the Bootstrap

Two dominant paradigms for statistics are the frequentist paradigm and the Bayesian paradigm. In each paradigm, we think of our sample data as drawn randomly from some population, such as US workers. Our goal is often to learn about some parameter that describes an aspect of this population, such as the average earnings of those workers.

In the frequentist paradigm, we treat the population as fixed but unknown, so that the only thing that is random is the sample we draw. Panel A of Figure 6 illustrates this idea. Here, we imagine a population of colors, with the composition of the population denoted by the area taken up by the color in the box on the left. A sample consists of a random set of colored circles, with the probability of each color determined by the population. The composition of the population itself is not actually known, but by observing a sample, we may be able to make inferences about the composition of the population. Intuitively, the more blue circles we see in our sample, the more consistent the sample is with a largely blue population.

In the Bayesian paradigm, we treat the population itself, in addition to the sample, as random. Before seeing the data, a Bayesian's uncertainty about the population is described by a *prior belief*. Panel B of Figure 6 illustrates this idea. Here, the prior belief specifies a distribution on the composition of the population. The true population is drawn (randomly) according to the prior belief, and then the observed sample is drawn randomly from the population. Because the sample is informative about the underlying population, after seeing the data, the Bayesian updates their view to form a *posterior belief*. Intuitively, the more blue circles appear

Figure 6
The Frequentist and Bayesian Paradigms for Statistics



Source: Authors' rendering.
Note: The diagrams illustrate the frequentist and the Bayesian paradigms for statistics. Panel A illustrates the frequentist paradigm, where the population is fixed (and unknown) and samples are randomly drawn from the population. Panel B illustrates the Bayesian paradigm, where both the population and the sample are random. In the Bayesian paradigm, we start with a prior belief, according to which many different populations can be randomly drawn. Then from any given population, we can randomly draw a sample.

in the sample, the more likely it is, under the posterior belief, that the population itself is mostly blue.

Because under the frequentist paradigm the population is fixed, we cannot talk coherently about the probability that the population is mostly blue, only about whether the sample data are consistent with a mostly blue population. We therefore often summarize uncertainty about the parameter of interest with a *confidence interval*. A confidence interval aims to contain the true value of the unknown parameter of interest with some prespecified probability, say 0.95, where the probability is with respect to the randomness of our sample.

By contrast, under the Bayesian paradigm, we can talk coherently about the *likely* contents of the population given the sample data. The Bayesian paradigm enables intuitive statements of the form, "Given the sample data, it is more likely than not that the population is majority blue." Under the Bayesian paradigm, we can summarize uncertainty about a parameter of interest with a *credible interval*. A

credible interval is an interval designed so that the posterior belief assigns some prespecified probability, say 0.95, to the true value of the parameter lying in the interval.

This takes us back to the bootstrap. Suppose we have run a nonparametric bootstrap. One way to construct a confidence interval is to compute the 2.5th and 97.5th quantiles of the bootstrap distribution. Under certain theoretical conditions, reviewed for example in Horowitz (2001, 2019), the resulting interval will contain the population value of the corresponding statistic 95 percent of the time—that is, in 95 out of 100 samples we might draw for our analysis. Under these conditions, then, we have constructed a valid frequentist confidence interval.[2]

In Example A at the start of the paper, the nonparametric bootstrap confidence interval for the treatment effect of calls on voting is $[0.032, 0.064]$. In the frequentist paradigm, we *cannot* say that the true treatment effect likely lies in this interval, but we *can* say that if we construct an interval in this way, then most of the time—that is, for most of the samples we could draw—the true treatment effect will lie inside the interval.

Of course, we can perform the same calculation using the Bayesian bootstrap. As it turns out, under very similar theoretical conditions (discussed, for example, in van der Vaart and Wellner 1996), the resulting interval will again contain the population value of the corresponding statistic 95 percent of the time and will thus be a valid frequentist confidence interval. In Example A, the Bayesian bootstrap confidence interval for the treatment effect of calls on voting is $[0.032, 0.064]$, very similar (identical up to rounding in fact!) to the interval we got from the nonparametric bootstrap.

However, when we use the Bayesian bootstrap, the resulting interval also has an interpretation as a credible interval anytime the observations are independent (and, as we discuss later, sometimes even when they are not). The reason is that the Bayesian bootstrap distribution is an approximation to a posterior belief. Specifically, the Bayesian bootstrap approximates the posterior belief of a Bayesian whose prior belief is of a type called a "Dirichlet process," formally introduced by Ferguson (1973, 1974), and whose prior has a small concentration parameter. Intuitively, this latter condition means that the posterior belief is heavily influenced by the sample data, which is why the Bayesian bootstrap distribution can be constructed entirely from the sample itself.

In Example A, the Bayesian bootstrap's credible interval for the cost per vote gained is $[\$6.6, \$13.1]$. Given this interval, and under the Bayesian paradigm, we *can* say that it is unlikely that the cost per vote gained is more than \$14.

In addition to enabling such intuitive statements, the Bayesian interpretation of the bootstrap can be powerful in other ways. In the next section, we turn to when

[2]These conditions typically involve large-sample approximations. Sometimes those approximations can be improved by constructing the bootstrap confidence interval in a different way, for example by nonparametrically bootstrapping a t-statistic (Horowitz 2019), or incorporating a bias correction (Efron 1987).

and how to use a bootstrap, where we will see more advantages of the Bayesian interpretation.

Recommendations for Practice

We turn now to some recommendations for use of the bootstrap in general and the Bayesian bootstrap in particular.

Run a Bayesian Bootstrap Whenever Possible

The bootstrap is a convenient and versatile tool for quantifying uncertainty. As we have seen, the Bayesian bootstrap is just as convenient as the nonparametric bootstrap, has some practical advantages, and can deliver both a confidence interval (under the frequentist paradigm) and a credible interval (under the Bayesian paradigm).

We therefore recommend running a Bayesian bootstrap whenever possible. Though there are formal principles for choosing the number of replicates (Hall 1986; Andrews and Buchinsky 2000), a simple rule of thumb is this. Divide your replicates in half. If the conclusions you'd reach from one half are substantively (economically) the same as those you'd reach from the other half, you probably have enough replicates (and if not, not).

Compare the Bayesian Bootstrap Distribution to the Normal Approximation

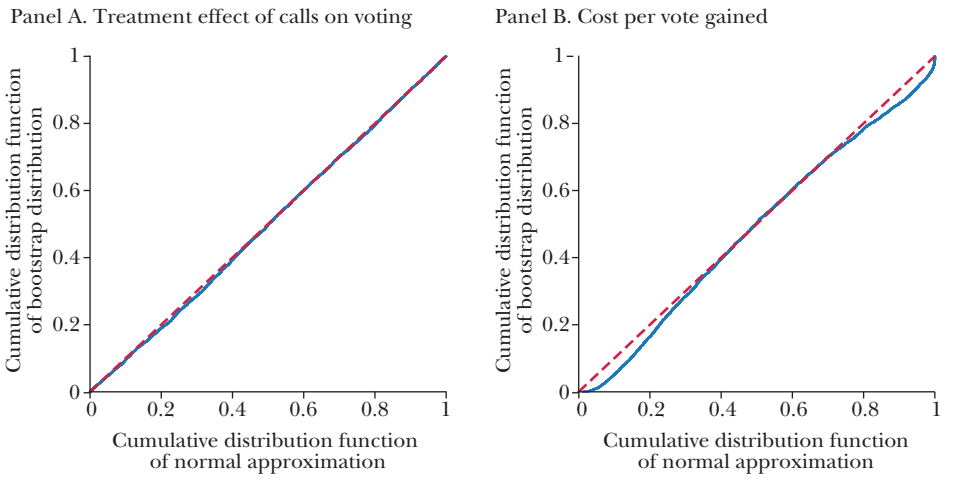
The common practice of reporting a point estimate and standard error can be justified by the large body of theoretical work showing that many statistics are approximately normally distributed in large samples. As we saw in Example A, however, such approximations can break down in realistic, and even simple, situations.

Fortunately, the bootstrap offers a generic tool for evaluating the quality of the normal approximation. Under many of the same kinds of theoretical conditions that guarantee a statistic is approximately normal, the bootstrap distribution will also be approximately normal. Thus, if the bootstrap distribution looks clearly different from the normal approximation, there is good reason to think that the normal approximation is unreliable. Angelini, Cavaliere, and Fanelli (2022, 2024) and Zhan (2026), among others, formalize this idea.

On the other hand, if the bootstrap distribution looks similar to the normal approximation, there is reason to think that the normal approximation is reliable, or at least reliable enough for many purposes. In particular, Andrews and Shapiro (2026) show conditions under which it is reasonable to use the normal approximation if and only if it looks similar to the bootstrap distribution.

To reach this conclusion, Andrews and Shapiro (2026) envision a researcher who reports their point estimate and standard error to a scientific audience including policymakers or other researchers. Members of the audience use the researcher's report to make decisions, for example setting policy or designing new experiments. Andrews and Shapiro (2026) ask when these decisions will be almost

Figure 7
Comparing the Bayesian Bootstrap Distribution to the Normal Approximation



Source: Calculations based on data from Nickerson, Friedrichs, and King (2006).
Note: The plots are p-p plots comparing the Bayesian bootstrap distribution to the normal approximation for the treatment effect of calls on voting (panel A) and the cost per vote gained (panel B). In both plots, the dashed line is the 45-degree line.

as good as those the audience would have made if they could have accessed the researcher's data directly. They find that, under certain conditions on the audience members' prior beliefs, and the decisions they need to make, this is true if and only if the normal distribution parameterized by the point estimate and standard error is similar to the bootstrap distribution.

Using their setup, Andrews and Shapiro (2026) also derive a straightforward measure of how different the normal approximation is from the bootstrap distribution. The measure can be read off what is called a "p-p plot," which compares two distributions by plotting their cumulative distribution functions against each other. If the two distributions are identical, their cumulative distribution functions will match and the plot will form a 45-degree line. The more the plot deviates from the 45-degree line, the more different the two distributions are. Andrews and Shapiro (2026) suggest measuring that difference by the sum of the largest positive and negative vertical distance between the plot and the 45-degree line (similar to a Kolmogorov-Smirnov statistic). The p-p plot is straightforward to make in standard software, and to make it even easier, Andrews and Shapiro (2026) created a web app that makes the p-p plot (and calculates the distance) from a point estimate, standard error, and bootstrap replicates.

Figure 7 displays p-p plots for the treatment effect and the cost per vote gained from Example A. Panel A shows that the normal approximation closely matches the bootstrap distribution for the treatment effect of calls on voting. Panel B shows

that, by contrast, the two distributions are meaningfully different for the cost per vote gained. In panel B, Andrews and Shapiro's (2026) measure of distance is large enough to be concerning. A campaign manager who relies on the normal approximation might make the wrong decisions about which methods to use for getting out the vote.

While the Andrews and Shapiro (2026) analysis can be extended to the nonparametric bootstrap, it applies most directly to the Bayesian bootstrap, which we think further strengthens the case for preferring the Bayesian bootstrap over the nonparametric bootstrap. We therefore recommend using the Bayesian bootstrap to evaluate the quality of the normal approximation. If, as in panel A of Figure 7, the normal approximation is pretty similar to the Bayesian bootstrap distribution, we think it is reasonable to report a point estimate and standard error.

Plot the Bayesian Bootstrap Distribution, or at Least Report Its Quantiles

What if, as in panel B of Figure 7, the normal approximation and the Bayesian bootstrap distribution look meaningfully different? In such a case, we have argued that the normal approximation is not a reliable guide to the uncertainty in the researcher's conclusions. Instead, we recommend reporting a plot, such as a histogram, of the entire Bayesian bootstrap distribution. Panel B of Figure 3 illustrates this recommendation by plotting a histogram of the Bayesian bootstrap distribution for the premium to leather tanning. The plot provides a rich picture of the Bayesian bootstrap distribution and avoids the risk of discarding useful information by providing too simple a summary.

What if the researcher's journal editor does not give the researcher as much space for plots as the *JEP* has generously given us? In that case, we recommend reporting, in addition to the point estimate, a credible interval formed by quantiles of the Bayesian bootstrap distribution: for example, a 95 percent credible interval formed by reporting the 2.5th and 97.5th quantiles of the bootstrap distribution. Such a credible interval will always respect the parameter's natural domain. For instance, in a case like Example A where the treatment effect is estimated as the difference between two proportions, such a credible interval will never imply that the treatment effect can be greater than one. Although the credible interval is an imperfect summary of the bootstrap distribution, we expect it will be useful in many situations, and researchers can always supplement it by including additional quantiles, or a plot of the bootstrap distribution, in an appendix.

Table 1 illustrates the reporting of the credible interval in the case of Example A. For both the treatment effect of calls on voting and the cost per vote gained, the table reports the point estimate and the credible interval implied by the Bayesian bootstrap. Notably, whereas the credible interval is symmetric about the point estimate for the treatment effect, it is not symmetric about the point estimate for the cost per vote gained. Intuitively, the possibility that the treatment effect may be small induces a substantial right tail in the Bayesian bootstrap distribution for the cost per vote gained, and this is reflected in the credible interval.

Table 1
Example Table Reporting a Credible Interval

Parameter	Point estimate	Bayesian bootstrap credible interval
Treatment effect of calls on voting	0.048	[0.032, 0.064]
Cost per vote gained	\$8.78	[\$6.59, \$13.12]

Source: Calculations based on data from Nickerson, Friedrichs, and King (2006).
Note: The table presents point estimates and Bayesian bootstrap credible intervals for the treatment effect of calls on voting and the cost per vote gained. The Bayesian bootstrap credible intervals are constructed by taking the 2.5th and 97.5th quantiles of the Bayesian bootstrap distribution.

Give the Bayesian Bootstrap a Bayesian Interpretation

The Bayesian interpretation of the Bayesian bootstrap makes a plot of the bootstrap distribution, such as the one in panel B of Figure 3, a powerful tool for communicating uncertainty to the researcher's audience. Recall that the Bayesian bootstrap distribution can be viewed as an (approximate) posterior belief about the parameter of interest for an agent with a particular type of prior. If the researcher's audience considers that type of prior to be reasonable, then the histogram summarizes a reasonable belief about the parameter of interest, given the researcher's data.

Importantly, the histogram can be useful even to an audience that doesn't think the type of prior underlying the Bayesian bootstrap is exactly right. Andrews and Shapiro (2026) formalize this conclusion using an old idea in Bayesian statistics, which is that if a researcher specifies a particular prior belief and reports the corresponding posterior belief to their audience, audience members whose prior beliefs differ from the one the researcher has specified can nevertheless learn a lot from the reported posterior. The reason is that audience members can adapt the reported posterior to reflect the differences between their own priors and the one the researcher has specified. An audience member who thinks the parameter of interest is likely to take on only very high values, for example, can adjust the reported posterior belief by putting more weight on high values and less weight on low values.

These ideas can justify reporting a bootstrap distribution as a kind of "benchmark" posterior belief for the audience. In this way, a histogram like the one in panel B of Figure 3 can be useful even to an audience that doesn't fully share the type of prior belief that underlies the Bayesian bootstrap.

Lastly, because the Bayesian interpretation of the Bayesian bootstrap does not rely on large-sample theory, none of the preceding interpretations requires accepting the validity of a normal approximation—or any other large-sample approximation—to the behavior of the statistic of interest.

For all of these reasons, we think that researchers should default to a Bayesian interpretation of the Bayesian bootstrap when writing and reading research articles, including (and perhaps especially) when the usual normal approximation breaks down.

If You Prefer a Frequentist Interpretation, Be Careful

For those who prefer a frequentist interpretation, it is important to note that a valid Bayesian credible interval is not necessarily a valid frequentist confidence interval. In situations where the usual normal approximation breaks down, or where the bootstrap does not well capture the usual normal approximation, bootstrap intervals may fail to contain the true value of the parameter with the desired probability (for example, 0.95).

In such cases, it may be possible to restore the validity of the confidence interval by using methods grounded in alternative approximations, depending on the particular reason why the normal or bootstrap approximation does not hold. In recent years, econometricians have developed such methods for a range of situations including weak instruments (Staiger and Stock 1997), parameters near a boundary (Andrews 2001), matching estimators (Abadie and Imbens 2008), highly nonlinear parameters (Andrews and Mikusheva 2016), directionally differentiable parameters (Fang and Santos 2019), and heavy-tailed distributions (Müller 2025). A number of such methods are in fact alternative bootstraps, such as subsampling (Politis and Romano 1994a), or the Gaussian multiplier bootstrap for high-dimensional settings (Chernozhukov, Chetverikov, and Kato 2013).

In our view, the fact that these approaches differ depending on the source of the failure of the normal or bootstrap approximation provides another reason to prefer the Bayesian interpretation of the Bayesian bootstrap, which applies generically to a wide range of situations, including most of those we list above.

Bootstrapping with Dependent Data

The version of the weighted bootstrap that we define above draws separate weights for each observation, implicitly treating the data as independent across observations. In practice, observations may be related to one another, for example due to panel structure. In this section, we discuss how to adapt the bootstrap to such settings.

Use a Bayesian Bootstrap with Clustering or Multiway Clustering

One common form of dependence is *clustering*, in which observations can be naturally grouped together in clusters: students within classrooms, workers within firms, households within villages, or individuals observed over time. Common analytical formulae for "clustered standard errors" implicitly treat the data as independent across clusters and dependent within clusters.

Encoding clustering in a weighted bootstrap is straightforward: Rather than drawing one weight per observation, we simply draw one weight per cluster, thus assigning the same weight to all observations in the same cluster. From there, the procedure is just as before: Calculate the statistic of interest on each reweighted sample and plot the resulting bootstrap distribution.

Another common form of dependence is *multiway clustering*, where a given observation can be a member of two (or more) different groups. For example, if we

observe data on wages for each of a set of workers at each of a set of firms at which they work, we may think the data are clustered *both* at the worker level and at the firm level.

The bootstrap readily adapts to such settings as well: We draw one weight per cluster on each dimension and then weight each observation by the *product* of all the weights for the clusters to which that the observation belongs. For example, a given worker-firm pair would get the weight given by the product of its worker's weight *and* its firm's weight.

In the case of the nonparametric bootstrap, we again use multinomial frequency weights, resulting in the *pigeonhole bootstrap* (Mccullagh 2000; Owen 2007; Davezies, D'Haultfœuille, and Guyonvarch 2021). In the case of the Bayesian bootstrap, we again use exponential weights, resulting in the procedure proposed in Sanders (2025). These procedures generally rest on an assumption that each observation (for example, worker-firm pair) behaves according to a function of the properties of each of its dimensions (that is, both the worker and the firm). Under such an assumption, the various practical and conceptual differences between the two bootstraps that we discuss above carry over directly to the multiway clustering setting.

Use a Different Bootstrap When the Data Are Not Clustered

So far, we've focused on settings where the data are independent or clustered, and we've argued that the Bayesian bootstrap is a good default in those situations. However, the data that economists work with can exhibit other forms of dependence for which the Bayesian bootstrap is not well-suited. In this section, we briefly discuss alternative bootstraps appropriate for some of these forms of dependence. Unlike the Bayesian bootstrap, these alternatives do not generally have a Bayesian interpretation, and the conditions under which they deliver valid frequentist inference vary.

With time series or spatial data, observations that are close in time or across space are related in ways that do not always create natural clusters. In some such situations, it is reasonable to suppose that the dependence between observations is small for observations that are far from one another in time or space. Such an assumption can justify *block-based bootstraps* (for example, the stationary block bootstrap of Politis and Romano 1994b), which resample blocks of observations. By keeping nearby observations together, block bootstraps preserve the dependence structure for close-by observations.

In situations with more complex dependencies, we may still be able to use a bootstrap if we are prepared to utilize the structure of an economic model. One such situation is where our economic model posits that the observed variables, which may be related in complex ways across observations, are determined by a set of fundamental shocks that are independent of one another and that enter the model as residuals. One way to take advantage of this structure is with a *residual bootstrap*. In a residual bootstrap, we first make a preliminary estimate of the parameters of our economic model. Using that estimate, we extract the implied residuals. In each

bootstrap replicate, we bootstrap the residuals, reconstitute the observed variables, and recompute our parameter estimate as if the reconstituted variables were the real data. A residual bootstrap can take different forms. One form resamples residuals with replacement, as in a nonparametric bootstrap (for example, section 7 of Efron 1979). Another form randomly changes the sign (and sometimes the magnitude) of the residual, resulting in what is often called a *wild bootstrap* (Wu 1986, Liu 1988, and Mammen 1993).

There are also situations in which we can reasonably resample neither clusters nor residuals, but in which our economic model is specified such that it is possible to simulate data from the model given a candidate value for the unknown parameter. In such a situation, we can apply the *parametric bootstrap* (Efron 1979 and Bickel and Freedman 1981). Here, we again start by estimating a preliminary value of the model's unknown parameters. In each bootstrap replicate, we simulate sample data from the model as if the preliminary estimate were correct. Then, we reestimate the parameters of the model using the simulated data. We think a parametric bootstrap is especially appealing in settings with complex dependencies (for example, through general equilibrium effects), which make alternative, model-agnostic bootstraps inapplicable.

The Bootstrap Will Not Do Economics for You

The Bayesian bootstrap is an extremely versatile and convenient procedure for quantifying statistical uncertainty. We conclude by noting that statistical uncertainty is not the only important form of uncertainty in the results that economists report.

Consider Example B from the start of the paper, focused on an industry earnings premium. There, the bootstrap is very helpful for describing the statistical uncertainty in a particular combination of regression coefficients, which together describe how a worker's earnings tend to change around the time the worker switches industries. The bootstrap does not, by itself, tell us anything about whether those wage changes are *caused* by industry switches: For example, those who moved from another industry to leather tanning might have seen their earnings fall even if they had not moved.

Instead, such interpretations rest upon economic assumptions, such as assumptions about the relationship between unobserved determinants of wages and industry moves. Because of the uncertainty that typically surrounds such assumptions, it would be ideal to have generic tools that allow researchers to quantify or communicate about that uncertainty. The ongoing quest for such tools is described in reviews by, for example, Andrews, Gentzkow, and Shapiro (2020), Armstrong (forthcoming), and Andrews, Chen, and Tecchio (forthcoming).

That said, statistical uncertainty is an important form of uncertainty. Thanks to the Bayesian bootstrap, it is one that is often straightforward to quantify. We recommend that you do so!

■ *We thank Peter Hull for contributions at the early stages of this project, and David Autor, Anna Bykhovskaya, Giuseppe Cavaliere, Bryan Graham, Phil Haile, Nicolas Jacquemet, Philipp Ketz, Jeffrey Kling, Xinwei Ma, Jonathan Parker, Steve Pischke, Will Rafey, Ashesh Rambachan, Jesse Rothstein, Andres Santos, Pietro Tebaldi, Elie Tamer, Timothy Taylor, Heidi Williams, and an anonymous reviewer for helpful comments. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under grant number 2141064. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.*

References

Abadie, Alberto, and Guido W. Imbens. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76 (6): 1537–57.

Andrews, Donald W. K. 2001. "Testing When a Parameter is on the Boundary of the Maintained Hypothesis." *Econometrica* 69 (3): 683–734.

Andrews, Donald W. K., and Moshe Buchinsky. 2000. "A Three-Step Method for Choosing the Number of Bootstrap Repetitions." *Econometrica* 68 (1): 23–51.

Andrews, Isaiah, Jiafeng Chen, and Otavio Tecchio. Forthcoming. "The Purpose of an Estimator is What it Does: Misspecification, Estimands, and Over-Identification." *Advances in Economics and Econometrics: Thirteenth World Congress*.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro. 2020. "Transparency in Structural Research." *Journal of Business and Economic Statistics* 38 (4): 711–22.

Andrews, Isaiah, and Anna Mikusheva. 2016. "A Geometric Approach to Nonlinear Econometric Models." *Econometrica* 84 (3): 1249–64.

Andrews, Isaiah, and Jesse M. Shapiro. 2026. "Communicating Scientific Uncertainty via Approximate Posteriors." *Econometrica* 94 (3): 843–75.

Andrews, Isaiah, Sarah Moon, and Jesse M. Shapiro. 2026. *Data and Code for*: "*In Praise of the Bayesian Bootstrap.*" Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E251784V1.

Angelini, Giovanni, Giuseppe Cavaliere, and Luca Fanelli. 2022. "Bootstrap Inference and Diagnostics in State Space Models: With Applications to Dynamic Macro Models." *Journal of Applied Econometrics* 37 (1): 3–22.

Angelini, Giovanni, Giuseppe Cavaliere, and Luca Fanelli. 2024. "An Identification and Testing Strategy for Proxy-SVARs With Weak Proxies." *Journal of Econometrics* 238 (2): 105604.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. 2017. "Leveraging Lotteries for School Value-Added: Testing and Estimation." *Quarterly Journal of Economics* 132 (2): 871–919.

Armstrong, Timothy B. Forthcoming. "Misspecification in Econometrics: A Selective Review." *Advances in Economics and Econometrics: Thirteenth World Congress*.

Bickel, Peter J., and David A. Freedman. 1981. "Some Asymptotic Theory for the Bootstrap." *Annals of Statistics* 9 (6): 1196–217.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2026. "A Users' Guide to Uncovering Worker and Firm Effects: The ABC of AKM." *Journal of Economic Perspectives* 40 (2): 195–214.

Chamberlain, Gary, and Guido W. Imbens. 2003. "Nonparametric Applications of Bayesian Inference." *Journal of Business and Economic Statistics* 21 (1): 12–18.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato. 2013. "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors." *Annals of Statistics* 41 (6): 2786–819.

Clyde, Merlise, and Herbert Lee. 2001. "Bagging and the Bayesian Bootstrap." In *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics* PMLR R3, edited by Thomas S. Richardson and Tommi S. Jaakkola, 57–62. Key West, FL: Proceedings of Machine Learning Research.

Davezies, Laurent, Xavier D'Haultfœuille, and Yannick Guyonvarch. 2021. "Empirical Process Results for Exchangeable Arrays." *Annals of Statistics* 49 (2): 845–62.

Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics* 7 (1): 1–26.

Efron, Bradley. 1987. "Better Bootstrap Confidence Intervals." *Journal of the American Statistical Association* 82 (397): 171–85.

Fang, Zheng, and Andres Santos. 2019. "Inference on Directionally Differentiable Functions." *Review of Economic Studies* 86 (1): 377–412.

Ferguson, Thomas S. 1973. "A Bayesian Analysis of Some Nonparametric Problems." *Annals of Statistics* 1 (2): 209–30.

Ferguson, Thomas S. 1974. "Prior Distributions on Spaces of Probability Measures." *Annals of Statistics* 2 (4): 615–29.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2021. "Place-Based Drivers of Mortality: Evidence from Migration." *American Economic Review* 111 (8): 2697–735.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Etienne Breton, and et al. 2025. *IPUMS CPS: Version 13.0 [Dataset]*. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D030.V13.0.

Ghosal, Subhashis, and Aad van der Vaart. 2017. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press.

Hall, Peter. 1986. "On the Number of Bootstrap Simulations Required to Construct a Confidence Interval." *Annals of Statistics* 14 (4): 1453–62.

Horowitz, Joel L. 2001. "The Bootstrap." In *Handbook of Econometrics*, Vol. 5, edited by James J. Heckman and Edward Leamer, 3159–228. Elsevier.

Horowitz, Joel L. 2019. "Bootstrap Methods in Econometrics." *Annual Review of Economics* 11 (1): 193–224.

Krueger, Alan B., and Lawrence H. Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica* 56 (2): 259–93.

Lei, Lihua, and Brad Ross. 2025. "Estimating Counterfactual Matrix Means With Short Panel Data." Preprint, arXiv. https://doi.org/10.48550/arXiv.2312.07520.

Liu, Regina Y. 1988. "Bootstrap Procedures under Some Non-I.I.D. Models." *Annals of Statistics* 16 (4): 1696–708.

Mammen, Enno. 1993. "Bootstrap and Wild Bootstrap for High Dimensional Linear Models." *Annals of Statistics* 21 (1): 255–85.

Mccullagh, Peter. 2000. "Resampling and Exchangeable Arrays." *Bernoulli* 6 (2): 285–301.

Müller, Ulrich K. 2025. "A More Robust t-Test." *Review of Economics and Statistics* 107 (3): 786–802.

Nickerson, David W., Ryan D. Friedrichs, and David C. King. 2006. "Partisan Mobilization Campaigns in the Field: Results from a Statewide Turnout Experiment in Michigan." *Political Research Quarterly* 59 (1): 85–97.

Owen, Art B. 2007. "The Pigeonhole Bootstrap." *Annals of Applied Statistics* 1 (2): 386–411.

Politis, Dimitris N., and Joseph P. Romano. 1994a. "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions." *Annals of Statistics* 22 (4): 2031–50.

Politis, Dimitris N., and Joseph P. Romano. 1994b. "The Stationary Bootstrap." *Journal of the American Statistical Association* 89 (428): 1303–13.

Rubin, Donald B. 1981. "The Bayesian Bootstrap." *Annals of Statistics* 9 (1): 130–34.

Sanders, Bas. 2025. "A New Bayesian Bootstrap for Quantitative Trade and Spatial Models." Preprint, arXiv. https://doi.org/10.48550/arXiv.2505.11967.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–86.

van der Vaart, Aad W., and Jon A. Wellner. 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Wu, C. F. J. 1986. "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis." *Annals of Statistics* 14 (4): 1261–95.

Xu, Li, Chris Gotwalt, Yili Hong, Caleb B. King, and William Q. Meeker. 2020. "Applications of the Fractional-Random-Weight Bootstrap." *American Statistician* 74 (4): 345–58.

Zhan, Zhaoguo. 2026. "Detecting Weak Identification by Bootstrap." *Econometric Reviews* 45 (7): 912–28.

Journal of Economic Perspectives—Volume 40, Number 4—Fall 2026—Pages 159–184

The Why and How of Foreign Aid

Dean Karlan

Foreign aid, whether privately or publicly funded and whether short-run humanitarian or long-run development, will not solve all problems yet receives many demands. How can resources be optimally deployed to achieve foreign aid goals and improve development outcomes? Consider two categories of questions for guidance: the why and the how.

The "why" includes: Why tackle a particular problem? Who benefits: the donor, the recipient, or the globe? To be in the national interest of the donor, is it sufficient to align with the will of the voters or must the aid benefit the donor via economic or geopolitical gains? Why address one problem at the opportunity cost of another? Policies must allocate resources across cause areas (health versus poverty), geographies (Guatemala versus Ghana), and time (immediate humanitarian needs versus reducing future suffering).

The "how" includes: How to tackle a particular problem? What is the market failure that underlies the problem, and what size of improvement can reasonably be expected? The "how" takes the "why" as given and drives the pragmatic decisions of whom to engage, what to fund, how to measure, how to decide all of the above, and how much impact to expect from proposed strategies. The "how" also circles back and informs the "why": If the "how" reveals a problem to be truly intractable, one

■ *Dean Karlan is the Frederic Esser Nemmers Distinguished Professor of Economics and Finance, Kellogg School of Management at Northwestern University, Evanston, Illinois. He served as Chief Economist, US Agency for International Development (USAID) from November 2022 to February 2025 and is the founder of Innovations for Poverty Action. His email address is dean.karlan@gmail.com.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251476.

may return to the "why" to redeploy resources toward a different, less intractable, problem.

In the language of economics, the "why" asks about allocative efficiency—given scarce resources and an abundance of problems, *cost-benefit* analysis guides choosing which problems to tackle. The "how" in contrast asks about productive efficiency—for a given problem, *cost-effectiveness* analysis guides choosing which strategy most efficiently addresses the problem identified in the "why" step.

Instability in the "why" can hinder development when progress requires long-term partnerships and continued investment. Motivations for publicly funded foreign aid swing with deliberate and informed rethinks, as well as erratic and unplanned winds (or gales, as experienced in 2025). While consensus on the "why" undoubtedly would yield stronger and more durable policy, disparate parties can often agree on a common plan despite different underlying motivations and incentives. For example, supporting a particular country's development effort could be motivated by donor self-interest in strengthening geopolitical power or moral impetus to reduce human suffering. Of course, when prescriptions differ—on the country in which to work, the problem to tackle, or the terms on which to act—then the political process must adjudicate.

The "how" drives whether the aid actually delivers on goals set by the "why." Evidence informing the "how" has improved radically in the past 30 years, with demonstrated impacts both within site (a program generates evidence that helps it iterate and improve) and via extrapolation (the upfront planning of a program is informed by evidence generated elsewhere). For the first, one World Bank (2025, p. 114) estimate of a "trial-and-adopt" iterative process concluded that program impact was often improved by over 50 percent from investing as little as 1 percent of the budget in testing. Returns to investment in evidence are even higher when lessons inform how resources elsewhere get allocated. The Massachusetts Institute of Technology Abdul Latif Jameel Poverty Action Lab and Innovations for Poverty Action report that evidence and policy advocacy from their networks have influenced programs and policies reaching over 850 million people (Dhaliwal 2025). For instance, school-based deworming programs that cost-effectively improve school attendance and future earnings have scaled to reach more than 280 million children (J-PAL 2022a); shifting from in-kind food distribution to electronic vouchers reached more than 18 million eligible households and reduced poverty in Indonesia at a lower cost to the government (J-PAL 2026); and an emissions market reduced air pollution at lower cost to firms and improved air quality for 15 million people in Gujarat, India (J-PAL 2025a). All were supported at some point in their design, testing, or scaling by foreign aid.

Beyond the direct benefits of improving how aid itself is used, three downstream benefits follow from improved evidence on what works and why. First, using aid to improve how low- and middle-income countries use *their own* resources can generate tremendous leverage, in both scale and durability of impact. If $10 million in technical support and funding of innovation to a $500 million ministry of health improves its effectiveness by 10 percent, that likely dwarfs what could have been

achieved by spending the $10 million on direct health services. Second, more effective programs could garner stronger political support, leading to a more stable and, thus, efficient system. Third, lower risk, all else equal, attracts more investment in addressing global problems.

Figure 1 illustrates the "why" and the "how" and their component steps as a decision-making flow chart. The chart inherently oversimplifies, skipping countless nuances and factors, but aims to capture several of the critical questions and how they influence key decisions such as when the "how" feeds back into the "why" and when to use existing evidence versus generate new evidence of effectiveness.

The Why of Foreign Aid: Multiple Motivations, Common Ground

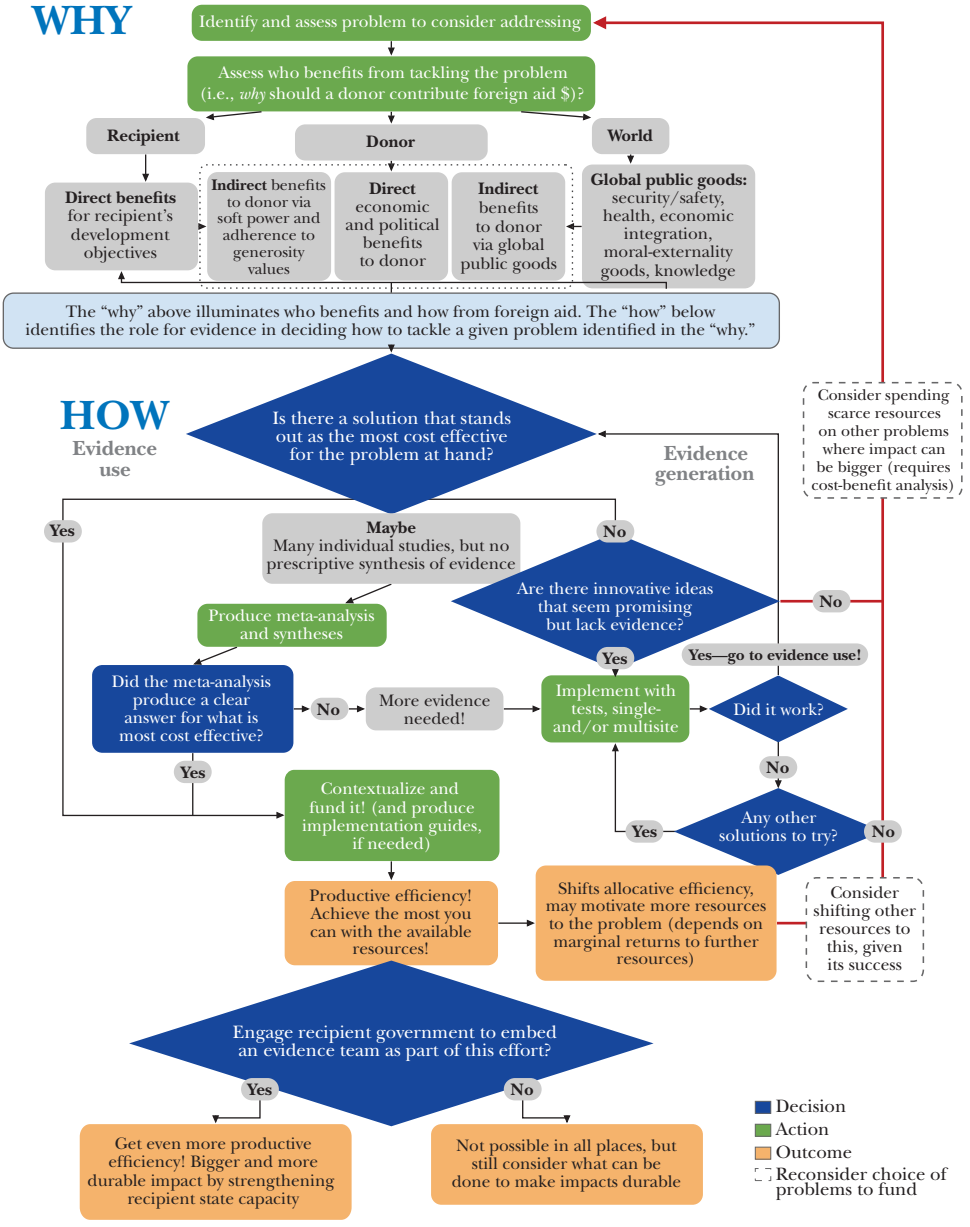
Most of the problems that aid seeks to address require long-term collaboration among many stakeholders. The challenge is to establish a "why" that is meaningful, logically coherent, and empirically supportable, yet broad enough to maintain strength and consistency through natural political ebbs and flows. This is not completely independent of the "how"; a logically coherent and empirically supportable "why" will consider the feasibility of tackling a particular problem.

Strikingly, donor-country political rhetoric consistently (particularly recently) invokes "national interest" as the primary justification for aid, yet leaves its definition vague. American officials have demanded that foreign aid make America safer, stronger, and more prosperous (Rubio 2025), while Germany's national security strategy states that development cooperation is both "in the interest of partner countries and also important for security in Germany" (Federal Ministry for Economic Cooperation and Development 2025). The United Kingdom's then-Secretary of State for International Development Andrew Mitchell (2011), in the same breath, called aid "morally right, it is part of what defines us as the British people… but it is also very much in our national interest."[1] This cross-national chorus reveals that "national interest" has become a phrase capacious enough to justify spending more as well as less. This blurring may not be mere rhetorical slippage, but rather a by-product of the actual overlap in the policies implied by these otherwise disparate motivations.

Categorizing the benefits of foreign aid into three types helps crystallize this conversation: benefits that accrue to (1) the donor, (2) the recipient, and (3) all via global public goods. Each dollar of foreign aid can serve multiple purposes, and the sum and the interactions across these benefits, weighted by the preferences of the

[1] Mitchell's (2011) framing echoes a much older domestic debate in Britain over whether state expenditure, rather than private charitable giving, better represents the authentic moral will of the citizenry. By the mid-twentieth century that debate had largely been settled in favor of the state: in a 1948 opinion poll, over 90 percent of Britons felt there was no longer any role for private charity (Breeze 2005). Prochaska (1988) traces this unfolding across the preceding century, as state provision slowly replaced Victorian-era voluntary philanthropy.

Figure 1
The Why and How of Foreign Aid



Source: Author's rendering.
Note: A simplified decisionmaking flowchart illustrating the component steps of the "why" and "how" of foreign aid.

decision-maker, drive foreign aid decision-making whether implicitly or explicitly (for a model illuminating this point, see Bermeo 2018, chapter 3).

Regardless of the "why," each path is strengthened when the aid itself is more effective at creating benefits for the recipients. A more meaningful gift is more appreciated and more likely to yield reciprocity. For example, effective service provision reduced local insurgent violence against the Iraqi government and Coalition forces in the Iraq war of 2003–2011 (Berman, Shapiro, and Felter 2011). Public opinion polls in 45 countries found that the United States' favorability improved after the launch of the President's Emergency Plan for AIDS Relief and the President's Malaria Initiative, visible and attributable programs found effective at achieving key outcomes (Jakubowski, Mai, Asch, and Bendavid 2019).

Direct Benefits: Economic and Political Returns to Donor Countries

The most direct form of donor benefit is an explicit quid pro quo: buying donor-country products to distribute in recipient countries or facilitating (through subsidy or technical support) trade and investment channels for donor-country firms to the recipient country. For example, the US Agency for International Development INVEST program, which provided support such as transaction advisory services and investment platforms, claimed to have mobilized $1.6 billion in private sector capital for development projects (Brown 2024). Claims of that sort—tallying up investments touched in some way by the consultants paid for by the aid contract—are commonplace in political accountability rhetoric, yet should not be confused as evidence of impact.

Beyond direct trade and investment, foreign aid serves what Bermeo (2017) terms "targeted development:" attempts by donor countries to limit negative spillovers from less-developed countries. Donor nations face challenges that cannot be optimally addressed through domestic spending alone: terrorism, unwanted migration, disease, trafficking in persons and illicit substances, and more. Here, the "quid" is an outcome in the recipient's national interest (for example, airport screening equipment) and the "quo" is an outcome in the donor's interest (for example, reduced terrorism risk). Bermeo shows that in the post-2001 era aid has become more targeted, increasingly aimed at achieving specific domestic priorities through spending abroad.

A direct quid-pro-quo aid effort can serve a domestic redistributive goal. For example, the US Food for Peace Title II program uses tax dollars to purchase American crops, ships those commodities largely via US-flagged vessels, and distributes them in countries experiencing food insecurity.[2] Foreign aid is likely an inefficient approach to domestic redistribution; a program that bought the same food and distributed it domestically would mechanically yield greater domestic benefits. Still, this form of aid could be one "why" behind foreign aid budgets and certainly is more salient and observable to the public than other benefits.

[2] The explicit aims were to "make maximum efficient use of surplus agricultural commodities," "stimulate and facilitate the expansion of foreign trade in agricultural commodities produced in the United States," and transfer "to any nation friendly to the United States in order to meet famine or other urgent relief requirements of such nation" (Agricultural Trade Development and Assistance Act, Pub. L. Nos. 83–480, 1954). When this aid was cut in 2025, many wrote about the potential negative impact on American farmers (Beras and Wong 2025).

Direct benefits need not be tied to the aid itself. In what is best described as hard power, the quid and the quo are immediate, specified, and likely coerced. The United States has often explicitly traded economic assistance for military advantage. For example, the United States forgave $7 billion in Egyptian debt as a reward for that country's participation in the Gulf War of 1990–1991 (Farnsworth 1991). In 2001, the United States waived sanctions on Pakistan and restarted economic assistance as a reward for cooperation in the war on terrorism (Wagner 2001). In 2009, the president of Yemen gave unrestricted permission for US unilateral strikes within the country against al-Qaeda; in 2010, security assistance to Yemen nearly tripled (Booth and Black 2010).

Indirect Benefits: Soft Power Returns to the Donor

The term "soft power" was coined by Joseph Nye, Jr. (1990, 2004) and defined as "the ability to get what you want through attraction rather than coercion or payment—magnets rather than carrots or sticks." A donor country may provide foreign aid expecting benefits that are not explicitly linked to the aid, such as an increase in its influence in global institutions or country-level engagement on specific markets or security matters. A cooperative equilibrium can arise, as in a repeated game in which reciprocity can be sustained without tying each individual act to a specific return.

The soft power rationale for foreign assistance has historical roots. For the United Kingdom, Hilton (2026) traces how humanitarianism became the primary vehicle through which Britain redefined its global role after decolonization, through both private philanthropy and nonprofits such as Oxfam, Save the Children, and Christian Aid, and through public funding. In the United States, President George W. Bush (2002) argued that "we fight against poverty because hope is an answer to terror," a formulation that includes a "targeted" pathway from aid to reduced poverty to greater enfranchisement to less terrorism, and a soft power pathway of increased influence and cooperation internationally.

While "soft power" is descriptively powerful, it can be analytically porous. The very breadth that makes "soft power" politically durable also makes it strategically incoherent, as it encompasses everything, yet commits to nothing specific. This porousness then provides political ammunition for aid expansions and contractions alike. The academic literature likewise can provide fodder either way, because soft power benefits are intrinsically indirect and diffuse and thus more vulnerable to both identification and generalizability challenges. This vulnerability is compounded by political pressure from both ends of the spectrum to bias findings in a desired direction, creating conditions in which "policy-based evidence-making" rules the discourse rather than evidence-based policymaking. Anecdotes, whether causal or not, can almost always be conjured up to make the case for or against foreign aid.

Altruistic Benefits: Generosity as Its Own Reward

Advanced democracies were shaped by ethical commitments such as religious social thought, human rights, and cosmopolitanism that favored the poorest people

over narrow self-interest (Lumsdaine 1993). Laboratory experimental evidence finds that giving itself generates happiness (Aknin, Mayraz, and Helliwell 2017). If the people of a donor country are motivated by some form of altruism, whether derived from philosophical principles or religious values, then a government engaging in humanitarian and development support is acting in its national interest, even if the aid generates no economic returns to the donor country.

Politicians of donor nations often speak of their nations as generous due to their great moral fiber; if so, aid for humanitarian reasons could strengthen community and national cohesion. The generosity motive thus yields returns for the donor nation by satisfying its citizens' stated desire to act altruistically and to be a nation that lives up to its own values (Yoho and Karlan 2025). Indeed, support for foreign aid in donor countries is generally high: for example, a 2014 Eurobarometer survey found that 73 percent of Europeans agreed that tackling poverty in developing countries is a moral obligation of the European Union (Eurobarometer 2015).

Global Public Goods: Benefits to Recipients, Donors, and the World

Under plausible assumptions, investing in global public goods can satisfy both private national interest and public good motivations. In particular, soft power gains are likely larger when the problem is global. If other donor countries follow suit, a cooperative equilibrium can emerge in which the public good is provided at a meaningful scale.

One set of foreign aid global public goods includes investments in reducing negative outcomes for others where transboundary threats pose potential economic or safety harm to donor countries too, such as climate change, contagious disease, forced displacement, war, and instability.[3] A second set comprises normative or moral-externality goods; the moral benefits of living in a world with less suffering are nonrival and nonexcludable. Examples include anti-human-trafficking programs and humanitarian responses to natural disasters. A third set encompasses knowledge public goods, which are typically nonrival and mostly nonexcludable, including scientific, technical, and policy knowledge on topics ranging from vaccines and agronomic innovations[4] to evidence on effective humanitarian and development policy.

This third category provides the conceptual bridge to the "how" of foreign aid. Development knowledge is nonrival: A rigorous evaluation demonstrating that a cash transfer program raises household consumption in Kenya is insightful to policymakers in Bangladesh, Bolivia, and Bihar. Regardless of which "why" motivates a particular donor, investing in the evidence base for what works promotes both national and global interests. The "how" is, therefore, not merely a technical

[3]Examples of tracking and response systems include the World Health Organization influenza work (WHO 2026) and the Intergovernmental Authority on Development's work on East Africa conflict (Intergovernmental Authority on Development 2026).

[4]An example is the Green Revolution, a period during which a series of agricultural innovations dramatically increased cereal output in low- and middle-income countries. The Consultative Group on International Agricultural Research, established in 1971, received multilateral and bilateral funding to create global centers for agronomic research (Byerlee and Lynam 2020).

afterthought to the "why;" it is itself a contribution to the global public good of development knowledge.

Before turning to the "how," a few notes on the sequencing of and relationship between the "why" and the "how." Grand strategies on the "why" established by donor-country leadership (like the executive branch leader, cabinet, and legislature) are often set with little regard for the "how." However, a level down, the "why" can be more responsive to the "how." In particular, if evidence on the "how" suggests that the best approach has a low probability of success, those setting the "why" may reconsider how they prioritized one cause area over another, one country over another, or one tool (for example, development or diplomacy) over another. For example, in 2022, the US government launched a "Democracy Delivers" initiative, which aimed to help young democracies strengthen their markets and institutions, with two broadly stated purposes: provide a global public good of strengthening democratic countries and provide direct benefits to the United States by facilitating trade between those countries and US firms (Blinken 2022). Determining the specific countries and cause areas was left to operating-level decision-makers that vetted ideas and, thus, effectively influenced the "why" as well as carried out the "how." Similarly, the US global aid response to the COVID-19 pandemic was defined as a priority by the highest levels of government and aimed to deliver humanitarian benefits globally (for example, by addressing extreme poverty and food insecurity) as well as direct economic and health benefits to the United States (by mitigating the pandemic's spread and economic fallout), but resources were allocated and later "surged" to countries according to criteria that prioritized need, capacity for success, and existing relationships (Dahl et al. 2024).

The How of Foreign Aid: The Key Role for Evidence

The "how" of foreign aid with respect to evidence can be organized around three interrelated strategies. The first is using existing evidence to choose and design programs, drawing on the accumulated body of rigorous impact evaluations to select interventions with a strong track record and adapt them to local context. The second, and most transformative, is scaling evidence-based policies through recipient governments or firms in low- and middle-income countries and strengthening their capacity to generate and apply rigorous evidence—not merely advising on one program at a time, but building the systems, culture, and institutional infrastructure for evidence-informed policymaking. The third is generating new evidence through single-site evaluations of innovations, coordinated multisite trials, or large does-it-work-at-scale studies. This step includes both impact assessments and finely tuned implementation research. These strategies intertwine: A strong evidence-use process not only identifies what evidence can and should be applied for scale, but also identifies the evidence that would reduce knowledge gaps and better inform future decision-making.

Two omissions are particularly worth noting. First, I omit many critical "how" questions that are outside the (narrowly construed) boundaries of economics,

such as how to legislate development cooperation, how to procure, how to choose funding recipients, how to structure bilateral versus multilateral engagement, and how to avoid fragmented bureaucracies (for three examples of many, see Burnside and Dollar 2000; Easterly, Levine, and Roodman 2004; Carcelli 2023). Second, the focus on cost-effectiveness evidence does not imply foreign aid should focus exclusively on programs that can be evaluated using experimental or quasi-experimental methods. Such methodologies have advantages when studying households, schools, villages, or firms, but are less applicable to many important development questions (for discussion, see Ravallion 2009). These include aid that operates at the country level, where credible comparison groups are difficult to construct; interventions whose effects unfold over decades; budget support or policy dialogue, where the "treatment" is diffuse and the counterfactual speculative; and many (but not all) governance and institutional reforms. This does not mean such aid cannot be evaluated: Natural experiments, modeling exercises, and qualitative case studies have roles to play. The leverage of such efforts can be huge in terms of resources eventually moved or efficiencies created through country-level or global reforms (for more on the role of state capacity, see Honig 2018; Muralidharan 2024). However, donor capacity is all the more important when undertaking aid efforts that are harder to design and measure.

The distinction between cost-benefit analysis and cost-effectiveness analysis is central here. Cost-benefit analysis facilitates adjudication of the "why": it weighs the expected monetized benefits of tackling one problem against another to determine where to allocate resources. Cost-effectiveness analysis guides the "how": Holding the problem constant, it asks which intervention, design, or delivery model produces the most impact per unit of expenditure. Two programs may generate similar impact per household in magnitude and outcome, but differ considerably in cost, thus rendering the cheaper program (all else equal) more cost-effective. The two are complementary but distinct. A program can be highly cost-effective yet still fail a cost-benefit test by solving (albeit efficiently) the less important problem. The "why" debate, adjudicated in part by cost-benefit reasoning, sets the agenda; the "how," guided by cost-effectiveness analysis, determines whether that agenda is actually delivered.

Cost-effectiveness analysis carries a practical advantage within the political economy of aid institutions. Foreign aid bureaucracies, like many large organizations, comprise constituencies loyal to specific problems such as maternal health, food security, climate adaptation, education, and others. Cost-benefit analysis, by asking whether a problem is worth tackling at all relative to alternatives, implicitly threatens those constituencies' raison d'être. Cost-effectiveness analysis is far less threatening: It accepts a group's goals and asks how to do the most good toward those goals with the resources in hand. Rather than asking a team to defend the very existence of their effort, cost-effectiveness evidence empowers them to achieve more with their resources. For example, a global panel of education experts ranked education interventions according to their relative cost-effectiveness at improving learning, giving donors and policymakers direction for what to implement (evidence use) as well as what to evaluate (evidence generation) (Global Education Evidence

Advisory Panel 2023). Thus, cost-effectiveness evidence is typically a more politically and bureaucratically viable path for improving aid. It also avoids the uncomfortable value-laden challenge of monetizing benefits, which economists embrace too readily as obvious and noneconomists decry as amoral (yet do so implicitly).

Building From the Evidence Base to Implementation at Scale

More than 20 years ago in this journal, Easterly (2003) asked whether foreign aid can buy growth and called for more evaluation and better incentives for both funders and recipients. Ten years later, Levine and Savedoff (2015) argued that building knowledge collectively is the future of aid. This period saw a large acceleration in the use of randomized impact evaluations, which was recognized in 2019 with a Nobel Prize in Economics awarded to Abhijit Banerjee, Esther Duflo, and Michael Kremer for their intellectual leadership in building this movement. A randomized evaluation, in its simplest form, randomizes the assignment process of an intervention or policy. The random process, when done with sufficient sample size such that the law of large numbers applies, creates treatment and control groups such that differences between them can be attributed causally to the intervention or policy, and not to outside factors, whether observable (such as weather or macroeconomic conditions) or difficult or impossible to observe (such as individual capabilities or preferences, which typically then could generate a selection bias). According to the International Initiative for Impact Evaluation website, there are now more than 8,000 randomized evaluations in development economics that shed light on the effectiveness of a variety of interventions and help to build external validity, where evaluations successfully identify mechanisms for why interventions work and where multiple evaluations can test for consistency with such theory across multiple contexts.

These efforts have enabled the development and spread of effective public policies reaching hundreds of millions of people. This plethora of evidence motivated the creation of the US Agency for International Development Office of the Chief Economist in 2023 (henceforth, OCE—an office I built and led from 2022 until resigning at the end of February 2025 when it, along with the rest of USAID, was dismantled). The OCE pushed for reform specifically to "use it or produce it"—that is, to either use evidence of cost-effectiveness to decide what to fund, or if the evidence was not deemed sufficient to be fully instructive, then to produce more evidence to better inform future decisions (USAID 2024).

"Use it or produce it" is the central fork of Figure 1: The question "Is there a solution that stands out as the most cost-effective for the problem at hand?" sends a "yes" down the evidence-use column on the left or a "no" down the evidence-generation column on the right, with a third, common, answer of "maybe" in the middle, when individual studies may exist but we lack sufficient evidence (and accompanying syntheses and aggregation) to confidently inform policymakers in a given context whether that solution is the best approach and how to adapt it.

The issue is not merely the quantity of evidence. Having an abundance of randomized evaluations is insufficient for tackling the obvious and necessary question: how does one apply evidence across space and time? How do we address

the "generalizability puzzle" of learning not merely "intervention X worked in context Y," but what about context Y was necessary to generate that result? Bates and Glennerster (2017) offer a framework for moving beyond vague "it depends, context matters" claims by instead clarifying which aspects of context matter for specific interventions so that the discussion does not stop with overly simplistic assertions that "country X is not country Y and, thus, nothing done in X can inform country Y policy." Ultimately, the goal is to combine theory that sheds light on when and how programs can work with data to test those theories (Deaton 2010).

Some new and creative policy innovations originate from nonprofits, some from governments of recipient countries, and some from academic researchers. Whatever the origin, testing and refining these innovations through randomized evaluations can shift the policy landscape. Here are three examples of programs that proceeded from local studies and pilot programs to a broad set of replicated studies and adaptations and adoptions at scale.

Conditional cash transfers, in which transfers are conditioned on school attendance, vaccinations, or health checkups, offer a first instructive example. In Mexico in the 1990s, and later in many other countries, governments worked with researchers to conduct randomized evaluations that consistently demonstrated that well-designed and contextualized conditional cash transfer programs can have positive impacts on health, food security, school attendance, consumption, and income. Subsequent research has learned that unconditional cash transfers have a similarly robust impact on a plethora of primary development objectives (Crosta et al. 2026). These results inspired governments around the world to create and test their own programs, both conditional and unconditional. As of 2016, over 60 governments worldwide have implemented a conditional cash transfer program and 130 have implemented unconditional cash transfers (Bastagli et al. 2016). Hundreds of rigorous evaluations have measured their impacts and helped governments learn what works to optimize program design and improve outcomes (Marliani and Pedro 2025). Recent evidence has gone further to study the macroeconomic effects of unconditional cash transfers, finding a macromultiplier effect of 2.5 (Egger et al. 2022), and efforts are now underway to expand and replicate this study in Malawi.

An education-related success story called Teaching at the Right Level (TaRL) offers a second example. The program builds foundational skills in math and reading by grouping children by learning levels, rather than teaching an identical curriculum to everyone at a given age or grade level. Pratham, one of India's largest education nongovernment organizations, developed TaRL in the early 2000s. Early proof-of-concept evaluations showed its effectiveness, and Pratham and others then tested iterations to understand the effectiveness of different delivery models (for example, implementation by village volunteers, nongovernment organization instructors, or government teachers). With strong repeated evidence of impact (Banerjee, Cole, Duflo, and Linden 2007; Angrist and Meager 2023), the program has now garnered attention from ministries of education for implementation at scale, with several related organizations (Pratham, Abdul Latif Jameel Poverty Action Lab, Youth Impact, and TaRL Africa) now providing technical support to the scaling entities (J-PAL 2022b).

A third example is the "Graduation" social protection approach, originally termed the "Graduation" model by the Bangladeshi nonprofit BRAC. The approach deploys a multifaceted livelihoods program that combines five complementary interventions (or some subset thereof), including a productive asset, training, coaching, savings access, and short-term consumption support, to provide a substantial push toward more secure and sustainable livelihoods for those living in extreme poverty. In 2017, the state government of Bihar in India began piloting a program to adopt the approach, to be implemented by a state agency with technical assistance from a nongovernment organization. The Bihar government subsequently announced plans to expand the Graduation approach in its "Satat Jeevikoparjan Yojana" (Sustainable Livelihoods Program), with the goal of extending the state's previous economic inclusion efforts to 200,000 of the poorest households across Bihar between 2018 and 2023, the first large-scale government adaptation of the Graduation approach (J-PAL 2025b).

Evidence from over 40 randomized controlled trials has shown that the Graduation approach for the very poor can increase income, consumption, productive assets, and other noneconomic outcomes such as mental health and community participation (for the initial randomized evaluations, see Banerjee et al. 2015; Bandiera et al. 2017; for syntheses, see Carter, Jee, Karlan, and Premand 2026; J-PAL 2025b). Ongoing evaluations are testing more specific implementation and design questions (such as which components are needed, how best to deliver the coaching and training, what the optimal bundle of transfers is, and so on). At present, BRAC and other technical service providers, typically nonprofit organizations with experience implementing the model, are working with governments in Ethiopia, India, Indonesia, the Philippines, Rwanda, and South Africa to integrate the model into existing systems at scale to achieve greater, more lasting impact. As of 2024, over 400 economic inclusion programs influenced by the Graduation approach have reached more than 15 million households (70 million people) globally (Partnership for Economic Inclusion 2024).[5]

Not all programs that demonstrate meaningful effects in pilots or in smaller-scale implementation succeed when implemented at scale. Banerjee et al. (2017) highlight examples of programs shown effective in a pilot that failed to deliver similar results at scale. Among these, a program that provided grants to hire extra teachers and reduce class sizes in Kenya raised student test scores when implemented by two different nongovernment organizations, but failed to do so when implemented at scale by the national government. The authors attributed this failure to weaker implementation capacity and adverse political circumstances. While some programs fail to demonstrate predicted impact at scale, others demonstrate greater impact via general equilibrium effects (for example, Egger et al. 2022; Muralidharan, Niehaus,

[5]The global health world is replete with almost countless examples of massive population health improvements stemming from evidence generated from randomized controlled trials. For a list of examples, the Center for Global Development has curated global health success stories at https://millionssaved.cgdev.org/.

and Sukhtankar 2023). The question "does it work at scale," in a sense, however, is a poorly posed question, as it bundles together disparate theoretical issues: organizational management and employee incentives, price effects, and political economy dynamics, among others.

In short, foreign aid can be spent to provide direct services, and there can be benefits from doing so, but when aid is combined with generation of new evidence, it can also influence future deployment of resources, whether public or private, in the recipient country or elsewhere. Generating knowledge—a public good—that identifies cost-effective approaches can further influence how other countries and donors deploy their resources, multiplying impact per dollar of aid and domestic resources alike (Karlan 2015).

Evidence Use and Generation: Strengthening Capacity of Aid Recipients with Evidence Partnerships

Foreign aid budgets may seem large, but are small relative to the budgets of the governments of low- and middle-income countries. Although governments want to implement effective policies, competing demands and limited resources hamper efforts to establish formal, systematic processes for applying and producing evidence. Governments need data to target, manage operations, and evaluate effectiveness of policies and programs, and on many other granular implementation details (akin to the "plumber" analogy in Duflo 2017). However, institutional frameworks that enable experimentation with program designs, pilot testing of new policy approaches, and integration of findings into decision-making are rare.

Formal collaborations with organizations committed to bridging the evidence-to-policy gap can help develop systems and processes that capitalize on this opportunity. The objective is not merely to partner on one research project or scale-up, but rather to change the culture of policymaking so that rigorous evidence becomes one of the key factors considered when choosing programs or policies. One approach is to embed teams dedicated to helping governments use and produce evidence directly within government agencies. Beyond implementing a single project, an embedded evidence lab identifies policy challenges, develops and implements strategies to generate relevant data and evidence, and facilitates the application of these insights. A lab can also strengthen an agency's ability to generate and use evidence by upgrading information systems for research and learning, building officials' capacity to demand and apply evidence, adapting institutional processes and legal frameworks to embed evidence-use practices, and securing the political and financial support needed for long-term sustainability. Here are three examples.

First, the Central Bank of Nigeria's Consumer Protection Embedded Evidence Lab was established in 2021 in partnership with Innovations for Poverty Action.[6] The lab sought to help the Central Bank of Nigeria integrate evidence-based decision-making into its mandate to safeguard financial consumers, who faced

[6]For an overview, see https://poverty-action.org/consumer-protection-nigeria-embedded-evidence-lab.

increasing risks as digital financial services expanded across the country. Nigeria's central bank grappled with a number of challenges in enforcing fair, transparent, and responsible conduct by financial institutions, including siloed consumer and financial data, lack of actionable analysis, and a rapidly evolving financial servicessector.

To address these challenges, the lab focused on improving the ability of the Central Bank of Nigeria to implement evidence-based regulatory policy through a four-pronged approach: (1) build capacity to apply data analysis tools in regulatory work; (2) upgrade administrative data systems to improve quality and reliability so that regulatory decisions are based on accurate and comprehensive information; (3) prioritize monitoring, evaluation, and learning to support decision-making and align with the bank's strategic objectives; and (4) create decision-oriented data products, including market monitoring reports, consumer complaint dashboards and institutional performance assessments. The collaboration improved the capacity of Nigeria's central bank to produce high-quality data, analytical tools, and reports for regulatory decision-making, all key to enhancing its oversight and accountability capabilities. The lab has become an integral part of the strategy of the Central Bank of Nigeria, contributing to multiple policy reforms and building a data-oriented culture.

Second, Peru's Environmental Evaluation and Enforcement Agency established an embedded evidence lab in 2024 to embed systematic data collection, rigorous impact evaluation, and iterative innovation into its regulatory and enforcement functions.[7] As the national authority responsible for balancing environmental protection with private investment, this agency seeks to deploy effective enforcement strategies consistently informed by empirical evidence. The lab uses a five-stage learning cycle: identify policy challenges, design and test solutions, analyze results, engage stakeholders to interpret results, and incorporate lessons into scaled-up approaches. Current work includes using satellite imagery to strengthen municipal compliance with waste management practices, testing interventions to increase use of a citizen reporting app, and developing a WhatsApp chatbot to boost adoption of an environmental emergency reporting tool.

Third, to improve the quality of learning in Rwanda, Innovations for Poverty Action and the Georgetown University Initiative on Innovation, Development and Evaluation partnered with Rwanda's Ministry of Education to conduct a randomized evaluation of a pay-for-performance teacher contracts program, called Supporting Teacher Achievement in Rwandan Schools (STARS), that led to improvements in both teacher effort and student learning (Leaver, Ozier, Serneels, and Zeitlin 2021). Rwanda already ranks highly among sub-Saharan African countries for education access; in 2015, net enrollment in primary education was 97 percent. However, improving education quality, reducing high rates of teacher turnover, and attracting teachers to rural areas remain critical challenges for the country's education sector.

[7]Established by presidential resolution (Resolution 00018-2025-OEFA/PCD 2025), the lab is formally housed within a unit responsible for developing innovative regulatory strategies and staffed by four government officials trained through structured capacity-building and learning-by-doing.

To build on these results, in 2019 the Ministry of Education established an embedded evidence lab to support the scale-up of the pay-for-performance approach and to iteratively test model refinements. With five full-time staff spread across the implementing agencies, the lab has direct access to policymaking processes, data systems, and implementation channels. The scale-up of the Supporting Teacher Achievement in Rwandan Schools program began in 248 schools, expanded to more than 340 schools, and is poised for a national scale-up covering 66,000 teachers and about 3.2 million students. The program includes strengthening data systems, training inspectors and headteachers, and embedding contract management within existing government structures to ensure sustainability. Early experimental results have shown positive treatment effects on teachers' lesson preparation, pedagogy, and consistent reporting of data to a ministry platform.

Scaling the Supporting Teacher Achievement in Rwandan Schools program took years, and securing political support was essential to the lab's sustainability. The lab maintained momentum by identifying small projects that contributed to the scaling effort while demonstrating its usefulness in real time—for instance, using testing data systems built for the randomized evaluation to inform short-term school management decisions on grade repetition and remedial tutoring. Beyond STARS, the lab increased demand for evidence more broadly and influenced institutional restructuring.

Evidence Generation: Innovation Funds

If aid effectiveness varies, and if innovation and evidence can improve it, there is a strong role for rigorous social science: generating evidence as a global public good, identifying cost-effective innovations, and further refining our understanding of effective approaches. This is the right-hand column of Figure 1: Promising ideas that lack evidence get implemented with tests, and the results either feed an implementation guide or send the decision-maker back to look for alternatives.

Development Innovation Ventures (DIV) at the US Agency for International Development demonstrated the potential for high returns from investing in development innovations as global public goods. The approach was broad, defining innovation to include new applications of technology as well as novel business models, delivery methods, production processes, products, or services that were expected to improve development outcomes. DIV also supported applied research and replications of proven innovations in novel environments and used a tiered model to fund three stages of innovation: piloting, testing for impact and cost-effectiveness, and transitioning innovations with rigorous evidence of impact and cost-effectiveness to scale. Like venture capital, most individual investments do not succeed, but a few high-impact innovations can generate exceptional returns that lift the entire portfolio. Kremer, Gallant, Rostapshova, and Thomas (2021) estimates a social benefit-cost ratio of at least 17:1 for the DIV portfolio between 2010 and 2019. The figure is a lower bound by construction: It compares the benefits of a subset of high-reach innovations against the cost of the entire portfolio, including everything that failed.

France's Fund for Innovation in Development, established in 2021, similarly employs a tiered funding model to support prototyping, rigorously evaluating already-prototyped ideas, and scaling those with strong evidence. So far, it has funded 98 projects in 30 countries, amounting to €50.1 million (Fund for Innovation in Development 2025).

Evidence Generation: Going Local, Globally, with Multisite Adaptations and Evaluations

Multisite impact evaluations—which evaluate interventions in multiple contexts with a common strategy—are another public good that foreign aid donors are uniquely positioned to fund. The Graduation approach was subject to a six-country multisite evaluation funded primarily by the Ford Foundation (Banerjee et al. 2015) and involved global coordination with some local adaptations (notable variations included the asset types and savings promotion methods). Multisite evaluations lay a foundation for a "built-in meta-analysis" that yields stronger lessons if coordination reduces unnecessary variation in design and measurement across sites, variation that would otherwise hinder interpreting cross-site treatment effect differences. The multisite design also allows for stronger inferences by expanding the context and populations for which the program is evaluated. Relative to uncoordinated studies, a multisite study can also reduce fixed costs, because sites can be added without having to duplicate all setup costs. Foreign aid donors, which typically fund similar programs in a diverse set of countries and often have in-house research teams, are well placed to identify, coordinate, and fund such studies and to use the results. For example, the World Food Programme and World Bank have conducted a number of multisite evaluations funded through thematic impact evaluation "windows"; among others, these include a four-country evaluation of cash transfers and gender impacts (DIME-WFP 2024) and a four-country study of resilience programs (World Food Programme 2019).

Multisite studies are not without challenges, including adapting interventions and identification and measurement strategies to each context and maintaining implementation fidelity to ensure comparability and accurate measurement. Still, when done well, this type of impact evaluation is a public good that holds significant potential to improve the effectiveness of foreign aid.

The evidence strategies described above—using existing knowledge, building government capacity, generating new rigorous evaluations—are tools for planning and allocating future resources more wisely. Often, budgets for evaluation are set by rough guidelines on a proportion of a program cost, but that is a mistake. Budgets for evaluation (and evidence generation more broadly) should be set by considering the size of the *future* budget that could be more informed and the likelihood that new evidence will inform that decision. Evaluation (and evidence generation more broadly) budgets should be larger when future budgets are larger, larger when evidence gaps are larger, smaller when evidence generation investments are less likely to shrink the knowledge gap, and smaller when political economy or other constraints are more likely to render the evidence unused.

Closing the Academic-Policy Gap: What Researchers Can Do Differently

The academic-policy gap regarding foreign aid is smaller than often claimed, but still too large. The thousands of randomized evaluations conducted in development over the past two decades have identified interventions with strong track records, and that evidence has been applied to scale programs and policies within and across countries—with cash transfers and Teaching at the Right Level programs offering two clear examples. Even so, for many topics, policymakers do not see evidence from elsewhere as prescriptive or helpful locally, and mismatched timelines and incentives are not always conducive to academics producing the outputs most useful to decision-makers. Two gaps are particularly consequential: insufficient synthesis and lack of practical contextualization guides.

The Synthesis Gap

While academics have produced a vast body of impact evidence, much of the knowledge remains fragmented. Systematic reviews and meta-analyses can steer funders toward more effective and efficient approaches by aggregating otherwise isolated evaluation findings into more actionable insights about what interventions and designs are most effective in which contexts. Importantly, evidence syntheses help identify knowledge gaps, which should then guide future allocation of resources to evidence generation and help policymakers know which programs and policies should undergo impact evaluations and which should only be monitored for implementation fidelity. In Figure 1, synthesis is one answer to the "maybe" branch, the common situation of many individual studies and no prescriptive synthesis, and whether the meta-analysis yields a clear answer on cost-effectiveness is the next question asked.

Despite substantial inroads in data sharing, continued effort is needed to close the synthesis gap. Aggregation has its challenges, such as requiring harmonization that may obscure key differences across studies (for example, testing the effect of business training by including a single indicator variable for "taught marketing"). Harmonization is difficult on three levels: outcome measurement, context descriptions, and program features. Selection bias in where and when programs are implemented, studied, and published creates further barriers. Even when data are pooled, careful interpretation is needed to understand whether, why, and how interventions work. More consistent reporting and harmonization of results, data, and measurement practices would facilitate the synthesis needed to identify approaches that consistently deliver results and make it easy for policymakers to take up and scale those "good buys."

Data about costs have improved considerably over the past 20 years, but important gaps remain. Often, evaluation papers focus on the treatment effects and underlying mechanisms for those effects, then render cost reporting to a simple per-recipient figure. An effort to report more granular cost data will help, but this too misses a key point: What matters most are the quantities of inputs and outputs of a program or intervention, far more than the cost of those inputs and outputs. For any policymaker who wants to decide what to do in *their* country, the prices

faced by the evaluated implementer in some other part of the world are far less relevant. Instead, the policymaker needs to know how many employees they will need to hire at what experience and background level in order to provide the desired program to the population they aim to serve. The policymaker can then use the wages in their market to determine the cost of implementing different options. Thus, academic evaluations should get better at reporting the production functions of programs in terms of the specific quantities of inputs and outputs, even more than the costs. Both, of course, would be ideal, so that more research can be done to understand (and thus improve) the production function for delivering products and services as part of public policies. Dioptra,[8] a consortium led by the International Rescue Committee, has made important inroads on this by building a platform for nine large international implementers to share detailed input, output, and price data. These data allow better modeling of production functions (to facilitate planning and cost-effectiveness analyses) and critical operational research that guides implementers on how to lower the cost of delivering the outputs to households, communities, or markets, holding quantity and quality fixed.

Furthermore, creating databases of study results would bridge a key gap in moving from evidence generation to aggregation, translation, and adaptation. Such efforts also would benefit the research process; for example, studies could include smaller control groups if the plethora of data from prior studies provided sufficient information to partially model the counterfactual. A recent collaboration between Northwestern University, the University of California, Berkeley, and Innovations for Poverty Action has launched "DEVIDENCE"[9]—a publicly available database of evidence from causal impact studies in international development—to provide structured and harmonized data on results from randomized evaluations to support evidence synthesis and meta-analyses.

Artificial intelligence (AI) can revolutionize the aggregation, harmonization, and synthesis of evidence, but at the same time, AI tools that deliver black box answers create risks. This guidance for using AI is not particular to policymakers, but also applies to the design of programs and interventions. For example, one chatbot tested in Kenya worked well for those with higher baseline business acumen, but created harm for others. Those with acumen had the judgement to know which advice to take, and those without did not (Otis et al. 2026). Used wisely, AI tools can radically lower cost and shorten time required to synthesize massive amounts of data and results. But if used without appropriate caution, poor policy choices may ensue. Policymakers can already pose questions to AI platforms such as "What is the best way to tackle such-and-such problem?" and AI will deliver an answer. Yet we do not have enough formal meta-analyses to serve as ground-truth data with which to then train and test whether the AI delivers reliable syntheses, nor have we curated sufficient underlying data from which to harmonize study results and generate bespoke quantitative AI-driven analyses and policy recommendations.

[8] For information about the Dioptra tool, see https://www.dioptratool.org/.
[9] For a description of DEVIDENCE, see https://devidence.org/.

Contextualization Gap and Implementation Guides

Adapting effective development programs to local contexts is crucial. Global evidence can point to cost-effective interventions, but failing to account for the local context when designing a program can be wasteful, or even harmful. For example, a financial planning board exercise for farmers in Zambia improved consumption smoothing during the hungry season: Farmers entered the hungry season with one additional month's food supply and achieved 9 percent higher crop revenue from their farms (Augenblick et al. 2026). However, when adapted to Malawi, the program needed to account for farmers' more diversified income streams as well as a scaling model that needed to work with groups rather than individuals to lower marginal costs (Center for Effective Global Action 2025).

Contextualization guides could also help alleviate an "untold" problem inherent in most academic papers. Papers report on the rationale for decisions that are central to the research question posed, like the specific elements that are randomized, varied, or tested. However, academic publications typically omit discussion of implementation decisions that were simply made and held constant: group sizes for interventions, participant recruitment methods, staff hiring processes, and other procedural choices. Such seemingly mundane decisions may influence the effectiveness of the program and its generalizability. Thus, a persistent gap arises between what is reported and what occurred, and between what is shared and what is usable by others. Academics could make it easier for policymakers to adopt evidence-based programs if implementation decisions were reported more thoroughly in accompanying documents.

Contextualization of evidence also requires balancing the "testing theories" imperative of research with less-obviously-theoretical research and/or merely descriptive policy needs. One way to address this is through foreign aid partnerships with local governments for embedded labs, described in the previous section. Breaking the cycle of single-experiment-to-single-paper studies would also enable more rapid-fire operational research to refine policy design. While this may be less rewarding academically than researching big theoretical questions, 100 relatively small and atheoretical tests, considered as a set, could reveal important policy and theoretical insights (for more on the case for operational research in development, see Angrist et al. 2026).

Taken together, closing the synthesis and contextualization gaps is not merely an academic exercise, but a public good in which philanthropists and foreign aid ought to invest (similar to the argument by Levine and Savedoff (2015) that building knowledge collectively is the future of aid). It is the mechanism by which all three strategies of the "how" reinforce one another: Existing evidence informs government action, government partnerships reveal gaps that motivate new evidence generation, and new evidence is most durable when produced by and for the institutions that will use it.

As one recent example of contextualization, the US Agency for International Development's Office of the Chief Economist incorporated guidance on implementing the Graduation approach into two $150 million resilience solicitations in

2024 (the year prior to USAID's dismantling). USAID had regularly made large, five-year awards for Resilience Food Security Activities. Historically, these USAID solicitations specified the outcome(s) to improve and asked applicants to propose interventions (often with some "illustrative examples" provided). Applicants were not asked to provide evidence of cost-effectiveness of their proposed interventions.[10] The lack of specificity, combined with a consensus approach to program design, led to a failure to face implicit-but-real tradeoffs, and thus produced what was pejoratively referred to as the "Christmas tree ornament" problem: Every constituency hung their favorite ornament on the program, whether it meshed well with the others or not.[11] At best, such lack of focus increases operational complexity and overhead costs; at worst, it reduces the resources allocated to the most effective components, hindering overall impact. The Office of the Chief Economist at USAID sought to combat this tendency by bringing forward the evidentiary base and steering program design either toward proven programs ("use it") or innovations alongside new evidence generation ("produce it"). With an embedded senior advisor on the design team, new solicitations became focused and evidence based, and the final plan would have larger impact and a broader reach (although all such awards were eventually cancelled).

As another example of contextualization, a large-scale randomized evaluation in Bangladesh helped identify an effective approach to increase mask-wearing during the COVID-19 pandemic: a community-level mask distribution and promotion program that tripled mask usage. The research team's support to implementing partners, including advising on procurement, locally contextualized communications materials, and a detailed implementation toolkit, enabled rapid scaling of the intervention to reach millions of people across South Asia. Abaluck et al. (2021) produced supplementary material for implementing a similar program, including contextualizing it for other settings. The materials clearly identify four key components of the program to normalize mask wearing and provide detailed guidance for procuring masks, preparing for fieldwork, and conducting surveillance (J-PAL 2021).[12]

This type of implementation and contextualization guidance facilitates rapid adoption and scaling of impactful interventions while also facilitating implementation fidelity and identifying key demand-driven knowledge gaps for further research to address.[13]

[10] As an example, a similarly sized program for Mozambique in 2022 issued such a solicitation (USAID 2021).

[11] Glennerster and Haria (2025) discusses the incentives that drive this complexity and how to address them to use aid budgets to best effect.

[12] For the full implementation toolkit, see https://drive.google.com/drive/folders/1QOcyQzxhXjPvrw1vUt4qVRT4Bisi6HZ.

[13] Another example comes from the What Works Hub for Global Education at https://www.wwhge.org/, with similar work on implementation guides for scaling structured pedagogy.

Conclusion

The "why" and "how" of foreign aid are too often treated as separate conversations—one belonging to diplomats and politicians, the other to program designers and evaluators. However, an ineffective "how" renders the "why" moot, and vice versa. The "why" and the "how" are also rarely as simple as they may sound: many statements are oddly broad or narrow depending on the unspoken fine print.

Of course, foreign aid is no magic wand for development. Aid can enable corruption, elite capture, and clientelism, which can spill over into overall governance (Svensson 2000). Governments may substitute aid for domestic revenue, weakening social contracts and undermining long-run capacity (Feyzioglu, Swaroop, and Zhu 1998; Benedek, Crivelli, Gupta, and Muthoora 2014). Large inflows of foreign currency from donors can appreciate the currency of the recipient country, similar to the way that the "resource curse" can appreciate the currency of commodity exporters, undermining other exports but also allowing for cheaper imports of often essential inputs such as fuel (Ahmad, Pentecost, and Stack 2023). Given the many factors that affect overall economic growth, it is difficult to make strong claims about how foreign aid affects recipient country macroeconomic growth rates (Rajan and Subramanian 2008). It is equally undeniable that many foreign aid efforts produced substantial benefits, with an abundance of evidence of individual programs producing critical, often life-saving, impacts.

Ultimately, as large as foreign aid budgets may seem, they remain small compared to government budgets of aid-recipient countries. The leverage point for sustained impact of aid is not the direct effect of a single demonstration program, but how the knowledge generated helps recipient governments develop the systems, political will, and technical capacity to demand and apply rigorous evidence for the betterment of their own policies and programs. Thus, aid donors should invest explicitly in evidence about foreign aid projects as a global public good—not primarily as an accountability requirement, but as a core strategy to improve future resource use. Also needed is more work on the transportability of evidence—that is, guidance for when lessons from Lilongwe apply to Lima, and when Jakarta in 2015 applies to Jakarta in 2030.

Three forms of humility are warranted. First, epistemic: "Does aid work?" is no more informative than "Does medicine work?" Both benefits and harms depend on the intervention and context. Second, institutional: Donors urging recipient governments to build state capacity should look in the mirror and improve their own capacity to design and deliver aid. Third, rhetorical: Grandiose claims may sound powerful, but lack proper underlying analysis. For example, a country remaining poor after decades of aid does not establish that aid failed; the relevant question is how it would have fared without that aid. Humility should deter us from false rhetoric for the sake of winning political points in public discourse.

What is not in serious doubt is that the tools now exist to do foreign aid dramatically better than it has historically been done. The evidence base is substantial, the methodology is mature, and the institutional models, from innovation funds to

embedded government labs, have proven their viability across diverse contexts. We must sharpen our demand even more for theory and evidence about what works, for whom, and under what conditions. Meanwhile, we have a lot of evidence to work with now and must work to close the gap between what the evidence says and what we do.

■ *I thank Sarah Bermeo, Kim Cernak, Deborah Cohen, Alison Fahey, Juan Manuel Hernández-Agramonte, Cindy Huang, Kelsey Jack, Gabriela Karlan, Chris Maloney, Håvard Nygård, Elina Scheja, Caitlin Tulloch, Anna Vanderkooy, and Christopher Udry for insightful feedback; Hope Cutchins, Elijah Goldberg, Megan McGuire, and Sneha Stephen for both insightful feedback and editing support; and, Nathan Ding and Vanessa Ikirezi for research assistance. All views and any misjudgments are my own.*

References

Abaluck, Jason, Laura H. Kwong, Ashley Styczynski, Ashraful Haque, Md. Alamgir Kabir, Ellen Bates-Jefferys, Emily Crawford et al. 2021. "Impact of Community Masking on COVID-19: A Cluster-Randomized Trial in Bangladesh." *Science* 375 (6577): eabi9069.

Ahmad, Ahmad Hassan, Eric J. Pentecost, and Marie M. Stack. 2023. "Foreign Aid, Debt Interest Repayments and Dutch Disease Effects in a Real Exchange Rate Model for African Countries." *Economic Modelling* 126: 106434.

Aknin, Lara B., Guy Mayraz, and John F. Helliwell. 2017. "The Emotional Consequences of Donation Opportunities." *Journal of Positive Psychology* 12 (2): 169–77.

Angrist, Noam, Tahir Andrabi, Stefan Dercon, Sharnic Djaker, James Habyarimana, Michelle Kaffenberger, Clare Leaver, Akshay Mangla, Nompumelelo Mohohlwane, and Abhijeet Singh. 2026. "An Agenda for Implementation Science in Education." *Science* 393 (6816): 1084–87.

Angrist, Noam, and Rachael Meager. 2023. "Implementation Matters: Generalizing Treatment Effects in Education." Annenberg Institute at Brown University EdWorkingPaper 23-802.

Augenblick, Ned, B. Kelsey Jack, Supreet Kaur, Felix Masiye, and Nicholas Swanson. 2026. "Retrieval Failures and Consumption Smoothing: A Field Experiment on Seasonal Poverty." NBER Working Paper 25430.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. "Labor Markets and Poverty in Village Economies." *Quarterly Journal of Economics* 132 (2): 811–70.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application." *Journal of Economic Perspectives* 31 (4): 73–102.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122 (3): 1235–64.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348 (6236): 1260799.

Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, Tanja Schmidt, and Luca Pellerano. 2016. *Cash Transfers: What Does the Evidence Say? A Rigorous Review of Programme Impact and of the Role of Design and Implementation Features.* Overseas Development Institute.

Bates, Mary Ann, and Rachel Glennerster. 2017. "The Generalizability Puzzle." *Stanford Social Innovation Review* 15 (3): 50–54.

Benedek, Dora, Ernesto Crivelli, Sanjeev Gupta, and Priscilla Muthoora. 2014. "Foreign Aid and Revenue: Still a Crowding-Out Effect?" *FinanzArchiv* 70 (1): 67–96.

Beras, Erika, and Wailin Wong, dirs. 2025. "Farmers Will Be Hit Hard by the Dismantling of USAID." *All Things Considered*. NPR, February 27. https://www.npr.org/2025/02/27/nx-s1-5311663/farmers-will-be-hit-hard-by-the-dismantling-of-usaid.

Berman, Eli, Jacob N. Shapiro, and Joseph H. Felter. 2011. "Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq." *Journal of Political Economy* 119 (4): 766–819.

Bermeo, Sarah Blodgett. 2017. "Aid Allocation and Targeted Development in an Increasingly Connected World." *International Organization* 71 (4): 735–66.

Bermeo, Sarah Blodgett. 2018. *Targeted Development: Industrialized Country Strategy in a Globalizing World*. Oxford University Press.

Blinken, Antony. 2022. "Secretary Antony J. Blinken at the USAID Democracy Delivers Event." Speech, USAID Democracy Delivers, New York, September 20. https://2021-2025.state.gov/secretary-antony-j-blinken-at-the-usaid-democracy-delivers-event/.

Booth, Robert, and Ian Black. 2010. "WikiLeaks Cables: Yemen Offered US 'Open Door' to Attack al-Qaida on Its Soil." *The Guardian*, December 3. https://www.theguardian.com/world/2010/dec/03/wikileaks-yemen-us-attack-al-qaida.

Breeze, Beth. 2005. "The Return of Philanthropy." *Prospect*, January 16. https://www.prospectmagazine.co.uk/essays/56671/the-return-of-philanthropy.

Brown, Nora. 2024. "So Long, Farewell: The End of INVEST." *Medium*, September 10. https://medium.com/usaid-invest/so-long-farewell-the-end-of-invest-e1c7df47b407.

Burnside, Craig, and David Dollar. 2000. "Aid, Policies, and Growth." *American Economic Review* 90 (4): 847–68.

Bush, George W. 2002. "Remarks to the United Nations Financing for Development Conference." Speech, UN Financing for Development Conference, Monterrey, Mexico. March 22. https://www.presidency.ucsb.edu/documents/remarks-the-united-nations-financing-for-development-conference-monterrey-mexico.

Byerlee, Derek, and John K. Lynam. 2020. "The Development of the International Center Model for Agricultural Research: A Prehistory of the CGIAR." *World Development* 135: 105080.

Carcelli, Shannon P. 2023. "Special Interests in Foreign Policy Bureaucracies: Evidence from Foreign Aid." *Journal of Politics* 85 (3): 905–18.

Carter, Samantha, Edward Jee, Dean Karlan, and Patrick Premand. 2026. "Doing Better at Doing Good: Lessons from a Bayesian Meta-Analysis after 20-Years of Poverty Graduation Programs." Unpublished.

Center for Effective Global Action. 2025. "Scaling up a Simple Planning Intervention to Help Farmers Save More of Their Harvest in Malawi." February 5. https://cega.berkeley.edu/feature/scaling-up-a-simple-planning-intervention-to-help-farmers-save-more-of-their-harvest-in-malawi/.

Crosta, Tommaso, Dean Karlan, Finley Ong, Julius Rüschenpöhler, and Christopher R. Udry. 2026. "Unconditional Cash Transfers: A Bayesian Meta-Analysis of Randomized Evaluations in Low and Middle Income Countries." NBER Working Paper 32779.

Dahl, Benjamin A., Beth Tritter, Deena Butryn, Melissa Dahlke, Sean Browning, Richard Gelting, Monica Fleming, et al. 2024. "Global VAX: A U.S. Contribution to Global COVID-19 Vaccination Efforts, 2021–2023." *Vaccine* 42 (S3): 125827.

Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48 (2): 424–55.

Dhaliwal, Iqbal. 2025. "The Evidence Effect: Ideas for the Future of Development." MIT Abdul Latif Jameel Poverty Action Lab (J-PAL) blog, May 12. https://www.povertyactionlab.org/blog/5-12-25/evidence-effect-ideas-future-development.

DIME-WFP. 2024. *Transforming Women's Participation via Cash-Based Transfer Programs*. World Bank and the World Food Programme.

Duflo, Esther. 2017. "The Economist as Plumber." *American Economic Review* 107 (5): 1–26.

Easterly, William. 2003. "Can Foreign Aid Buy Growth?" *Journal of Economic Perspectives* 17 (3): 23–48.

Easterly, William, Ross Levine, and David Roodman. 2004. "Aid, Policies, and Growth: Comment." *American Economic Review* 94 (3): 774–80.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker. 2022. "General Equilibrium Effects of Cash Transfers: Experimental Evidence From Kenya." *Econometrica* 90 (6): 2603–43.

Eurobarometer. 2015. "The European Year for Development—Citizens Views on Development, Cooperation and Aid." European Commission, January. https://europa.eu/eurobarometer/surveys/detail/2022.

Farnsworth, Clyde H. 1991. "Egypt's 'Reward': Forgiven Debt." *New York Times*, April 10. https://www.nytimes.com/1991/04/10/business/egypt-s-reward-forgiven-debt.html.

Federal Ministry for Economic Cooperation and Development. 2025. "The German Government's National Security Strategy." German Federal Ministry for Economic Cooperation and Development, June 18. https://www.bmz.de/en/ministry/german-national-security-strategy.

Feyzioglu, Tarhan, Vinaya Swaroop, and Min Zhu. 1998. "A Panel Data Analysis of the Fungibility of Foreign Aid." *World Bank Economic Review* 12 (1): 29–58.

Fund for Innovation in Development. 2025. "FID at a Glance." Fund for Innovation in Development. https://fundinnovation.dev/en/impact#discover.

Glennerster, Rachel, and Siddhartha Haria. 2025. "Radical Simplification: A Practical Way to Get More Out of Limited Foreign Assistance Budgets." Center For Global Development, April 21. https://www.cgdev.org/blog/radical-simplification-practical-way-get-more-out-limited-foreign-assistance-budgets.

Global Education Evidence Advisory Panel. 2023. *Cost-Effective Approaches to Improve Global Learning: What Does Recent Evidence Tell Us Are "Smart Buys" for Improving Learning in Low- and Middle-Income Countries?* World Bank.

Hilton, Matthew. 2026. *Charity after Empire: British Humanitarianism, Decolonisation and Development.* Cambridge University Press.

Honig, Dan. 2018. *Navigation by Judgment: Why and When Top-Down Management of Foreign Aid Doesn't Work.* Oxford University Press.

Intergovernmental Authority on Development. 2026. "IGAD's Conflict Early Warning and Response Mechanism (CEWARN)." IGAD Djibouti. https://igad.int/cewarn-2/.

Jakubowski, Aleksandra, Don Mai, Steven M. Asch, and Eran Bendavid. 2019. "Impact of Health Aid Investments on Public Opinion of the United States: Analysis of Global Attitude Surveys From 45 Countries, 2002–2016." *American Journal of Public Health* 109 (7): 1034–41.

J-PAL. 2021. "Normalizing Mask Usage to Combat Covid-19." October 27. https://www.povertyactionlab.org/case-study/normalizing-mask-usage-combat-covid-19.

J-PAL. 2022a. "Deworming to Increase School Attendance." September 22. https://www.povertyactionlab.org/case-study/deworming-increase-school-attendance.

J-PAL. 2022b. "Teaching at the Right Level to Improve Learning." August 11. https://www.povertyactionlab.org/case-study/teaching-right-level-improve-learning.

J-PAL. 2025a. "Emissions Markets to Reduce Air Pollution." July 1. https://www.povertyactionlab.org/case-study/emissions-markets-reduce-air-pollution.

J-PAL. 2025b. "Targeting the Ultra-Poor to Improve Livelihoods." May 14. https://www.povertyactionlab.org/case-study/targeting-ultra-poor-improve-livelihoods.

J-PAL. 2026. "Digitizing Social Services to Reach the Right People and Reduce Poverty." January 13. https://www.povertyactionlab.org/case-study/digitizing-social-services-reach-right-people-and-reduce-poverty.

Karlan, Dean. 2015. "The Multi-Lateral Development Banks: A Strategy for Generating Increased Return on Investment." Testimony before the US House Committee on Financial Services.

Kremer, Michael, Sasha Gallant, Olga Rostapshova, and Milan Thomas. 2021. "Is Development Economics a Good Investment? Evidence on Scaling Rate and Social Returns from USAID's Innovation Fund." Unpublished.

Leaver, Clare, Owen Ozier, Pieter Serneels, and Andrew Zeitlin. 2021. "Recruitment, Effort, and Retention Effects of Performance Contracts for Civil Servants: Experimental Evidence from Rwandan Primary Schools." *American Economic Review* 111 (7): 2213–46.

Levine, Ruth, and William Savedoff. 2015. "The Future of Aid: Building Knowledge Collectively." Center for Global Development Policy Paper 050.

Lumsdaine, David Halloran. 1993. *Moral Vision in International Politics: The Foreign Aid Regime, 1949–1989.* Princeton University Press.

Marliani, Lina, and Paula Pedro. 2025. "The Evidence Effect: How LMIC Governments Drive and Scale Effective Poverty Reduction Policies." J-PAL blog, June 5. https://www.povertyactionlab.org/blog/6-5-25/evidence-effect-how-lmic-governments-drive-and-scale-effective-poverty-reduction.

Mitchell, Andrew. 2011. "Interview on Foreign Aid." *BBC News*, March 1. https://www.bbc.com/news/av/uk-politics-12606763.

Muralidharan, Karthik. 2024. *Accelerating India's Development: A State-Led Roadmap for Effective Governance.* Penguin Random House India.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2023. "General Equilibrium Effects of (Improving) Public Employment Programs: Experimental Evidence from India." *Econometrica* 91 (4): 1261–95.

Nye, Joseph S., Jr. 1990. *Bound to Lead: The Changing Nature of American Power.* Basic Books.

Nye, Joseph S., Jr. 2004. *Soft Power: The Means to Success in World Politics.* PublicAffairs.

Otis, Nicholas G., Rowan Clarke, Solène Delecourt, David Holtz, and Rembrand Koning. 2026. "The Uneven Impact of Generative Artificial Intelligence on Entrepreneurial Performance: Evidence from a Field Experiment in Kenya." *Management Science.* https://doi.org/10.1287/mnsc.2024.06909.

Partnership for Economic Inclusion. 2024. *2024 Annual Report.* World Bank.

Prochaska, Frank. 1988. *The Voluntary Impulse: Philanthropy in Modern Britain*. Faber and Faber.

Rajan, Raghuram G., and Arvind Subramanian. 2008. "Aid and Growth: What Does the Cross-Country Evidence Really Show?" *Review of Economics and Statistics* 90 (4): 643–65.

Ravallion, Martin. 2009. "Evaluation in the Practice of Development." *World Bank Research Observer* 24 (1): 29–53.

Rubio, Marco. 2025. "Opening Statement of Marco Rubio to Be Secretary of State." Speech, US Senate Committee on Foreign Relations, Washington, DC, January 15. https://www.foreign.senate.gov/imo/media/doc/6df93f4b-a83c-89ac-0fac-9b586715afd8/011525_Transcript_Nomination%20of%20Marco%20Rubio%20to%20be%20Secretary%20of%20State.pdf.

Svensson, Jakob. 2000. "Foreign Aid and Rent-Seeking." *Journal of International Economics* 51 (2): 437–61.

USAID. 2021. "Request for Applications for a Resilience Food Security Activity in Mozambique." Funding Opportunity Number: 720BHA21RFA00001. https://files.simpler.grants.gov/opportunities/18ffb253-f06b-4906-aff3-9341f7d1ddbb/attachments/6f78adc3-bef2-4112-97ba-16028ca4f6c9/FINAL_Mozambique_FY22_RFSA_RFA.pdf.

USAID. 2024. USAID Position Paper: Cost Effectiveness. USAID.

Wagner, Alex. 2001. "Bush Authorized to Lift Sanctions on Pakistan." Arms Control Association. https://www.armscontrol.org/act/2001-11/press-releases/bush-authorized-lift-sanctions-pakistan.

World Food Programme. 2019. "Resilience in a Changing Climate: Impact Evaluation Window." World Food Programme.

WHO. 2026. "Global Influenza Surveillance and Response System (GISRS)." https://www.who.int/initiatives/global-influenza-surveillance-and-response-system.

World Bank. 2025. *Driving Global Impact: Accelerating Economic Development with Trial-and-Adopt, LEADS, and ImpactAI.* World Bank.

Yoho, Ted, and Dean Karlan. 2025. "A Theory and Evidence Test for Strategic International Aid." *The Hill*, March 27. https://thehill.com/opinion/international/5215080-reimagined-foreign-aid-strategy/.

Journal of Economic Perspectives—Volume 40, Number 4—Fall 2026—Pages 185–206

The Role of Organization in Firm-Sponsored Training

David J. Deming, Raffaella Sadun, and Christopher Stanton

Some firms invest heavily in worker development, while other observationally similar competitors invest almost nothing. Why do otherwise comparable firms make such different choices about who to train, how much to invest, and what skills to build?

These differences matter. If wage growth reflects rising productivity, most life cycle human capital accumulation happens at work rather than in school (for example, Deming 2025). A growing body of evidence on firm effects points to the importance of early career employment in high-wage-growth firms as a pronounced source of lifetime earnings variation. Even within the same industry and labor market, high-wage-growth firms offer distinct early career wage growth paths that persist for decades (Di Addario, Kline, Saggio, and Sølvsten 2023; Arellano-Bover and Saltiel 2026). Theories of within-firm learning and task assignment (Gibbons and Waldman 2006) and structural analysis of the returns to training (Adda and Dustmann 2023) point to firm-sponsored skill investment as a leading candidate mechanism for these differences, though direct evidence remains thin.

Standard labor economic theories of on-the-job training start from Becker's (1964) famous argument that firms in perfectly competitive labor markets will not pay for general human capital because the benefits accrue only to workers who are

■ *David J. Deming is Professor of Economics and Public Policy, Harvard University, Cambridge, Massachusetts. Raffaella Sadun and Christopher Stanton are both Professors of Business Administration, Harvard Business School, Boston, Massachusetts. Their email addresses are david_deming@harvard.edu, rsadun@hbs.edu, and cstanton@hbs.edu. Author names are listed in alphabetical order.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20261508.

being paid the marginal product of their labor. Subsequent work has described labor market frictions that can allow firms to appropriate part of the surplus from training investments: As one example, if a firm has monopsony power in the labor market or if employees face search frictions, wages are compressed relative to productivity (as in Acemoglu and Pischke 1999), and firm-sponsored training even for general human capital can be viable when labor market imperfections allow firms to capture some of the returns from firm outlays on training.

A different account of these facts starts from firms choosing to train workers as a response to organizational complementarities that increase the overall returns to skill investment, not only the share of those returns a firm can capture. Under this view, training is embedded in a broader bundle of productivity-enhancing practices (performance management, promotion ladders, and technology adoption) that vary systematically across firms. Some firms provide intensive worker development at entry and little thereafter; others invest continuously in frontline upskilling; still others concentrate resources on managerial and leadership training higher in the hierarchy. The result is an employment landscape in which learning opportunities—and therefore human capital accumulation—are deeply shaped by organizational context. This perspective accounts for the within-industry variation in training across firms and shifts the question from *why firms train* to *why some firms build organizations in which training is valuable.*

We begin by characterizing some empirical patterns in on-the-job training, but in this task, we confront a lack of systematic data on training. The US government has not carried out a systematic survey of workplace training since the 1995 Survey of Employer-Provided Training. The only available US data on spending levels for on-the-job training is from industry sources not designed for research use. The limited available evidence suggests that US firms spend on the order of $100 billion a year on training, with expenditures reaching roughly 4–5 percent of labor costs in the 1990s (Black, Skipper, and Smith 2023); in a nationally representative 2022 survey of employed US workers, 87 percent reported that their employer offers new-hire training and 80 percent reported ongoing training or continuing education (Sandvik, Saouma, Seegert, and Stanton 2026).

However, there are several sources of survey data concerning which workers within an organization received on-the-job training. We draw upon several of these surveys to document four empirical patterns. First, participation in training rises with workers' education, suggesting complementarity between schooling and on-the-job learning. Second, firm characteristics—such as industry of activity, establishment size, and learning culture—explain a large fraction of the variation in participation in formal and informal training activity across workers, even after controlling for individual and job characteristics. Third, better-managed firms invest more in training and do so in more sophisticated forms. Fourth, training strategies cluster: Some firms emphasize managerial development and internal promotion ladders, while others focus on standardized frontline skills. These patterns place firm-level capabilities at the center of adult skill formation.

What makes some organizations more valuable training contexts than others? We argue that scarce managerial talent is the key driver of the variation in organizational design choices and training investments. In the "knowledge hierarchies" framework of Garicano (2000) and Garicano and Rossi-Hansberg (2006), higher-level managers spend their limited time solving the most complex problems and lower-level managers address more common tasks. In this way, the productivity of the top of the management hierarchy depends on the filtering and problem-solving capacity of the middle and lower layers. In this framework, training not only increases worker productivity directly, but also frees time higher up in the hierarchy, increasing managerial effectiveness and/or spans of control (Espinosa and Stanton 2026). As a result, we expect firms with stronger managerial capability to invest more in training, and the level of training should cluster with other organizational practices.

This perspective has implications for wage inequality and corresponding public policy. If firms differ in their capacity to generate and deploy human capital, then early career sorting into high- or low-capability firms can produce persistent wage divergence, even among observationally similar workers. Moreover, policies that target individual barriers to training—such as tuition subsidies to help overcome credit constraints—may have limited effects if firm capability is the binding constraint for deployment of learning programs for employees. Improving human capital accumulation at scale may require attention to management quality, organizational systems, and complementarities between training and production.

How does the incentive to hire and train junior workers change with recent advances in generative artificial intelligence (AI)? The job market for entry-level workers has been weak in the years since ChatGPT was released, with relatively greater impacts in AI-exposed occupations (for example, Brynjolfsson, Chandar, and Chen 2025). Some authors have concluded that AI represents seniority-biased technological change that disproportionately reduces demand for junior workers (Hosseini Maasoum and Lichtinger 2025). Garicano and Rayo (2026) develop a model that explores how AI changes training incentives in a knowledge hierarchy. In their model, AI disrupts the implicit bargain in which junior workers "pay" for training by providing rote, entry-level work. If AI is a better substitute for junior work, firms have less of an incentive to train. In our framework, the value of firm training also depends on organization-specific context, which could make training valuable even as AI capabilities advance.

Throughout this essay, we will use "training" and "workplace learning" somewhat interchangeably, but the distinction matters and deserves acknowledgment. Formal training programs with structured curricula and credentialing are only one mechanism through which firms build human capital. Peer learning, mentorship, job rotation, and the day-to-day accumulation of know-how through experience or trial and error are likely just as important, but available data rarely permit a clean separation. We expect formal training programs and the informal learning that surrounds them to be strongly complementary, particularly where firms structure formal programs deliberately around who participates, whether it is mandatory, and how training connects to promotion.

Who Gets Training—and How Firms Organize It

Becker's (1964) baseline model of human capital accumulation, when applied to competitive labor markets, suggests that any firm-sponsored general skills training is paid for in some way by the employee, most likely through lower wages. Yet there appears to only be a modest or weak relationship between training investments and lower wages in analyses dating back to data from 1992 in the United States (Barron, Berger, and Black 1999). More recent analyses of workers in Brazil and Italy show that employees of firms that generate high lifetime earnings, plausibly through training, do not start at lower wage levels (Arellano-Bover and Saltiel 2026). Instead, productivity gains that come from training partially accrue to the employer but also appear to benefit workers, raising the question of which workers receive these opportunities.

This section uses three complementary data sources to explore this question. The Programme for the International Assessment of Adult Competencies (PIAAC) identifies how training participation varies across workers and firms in 29 countries. The World Management Survey links training provision to management practices in a large sample of manufacturing firms. A 2023 survey of chief human resources officers at US firms lets us look at how training is organized inside firms.

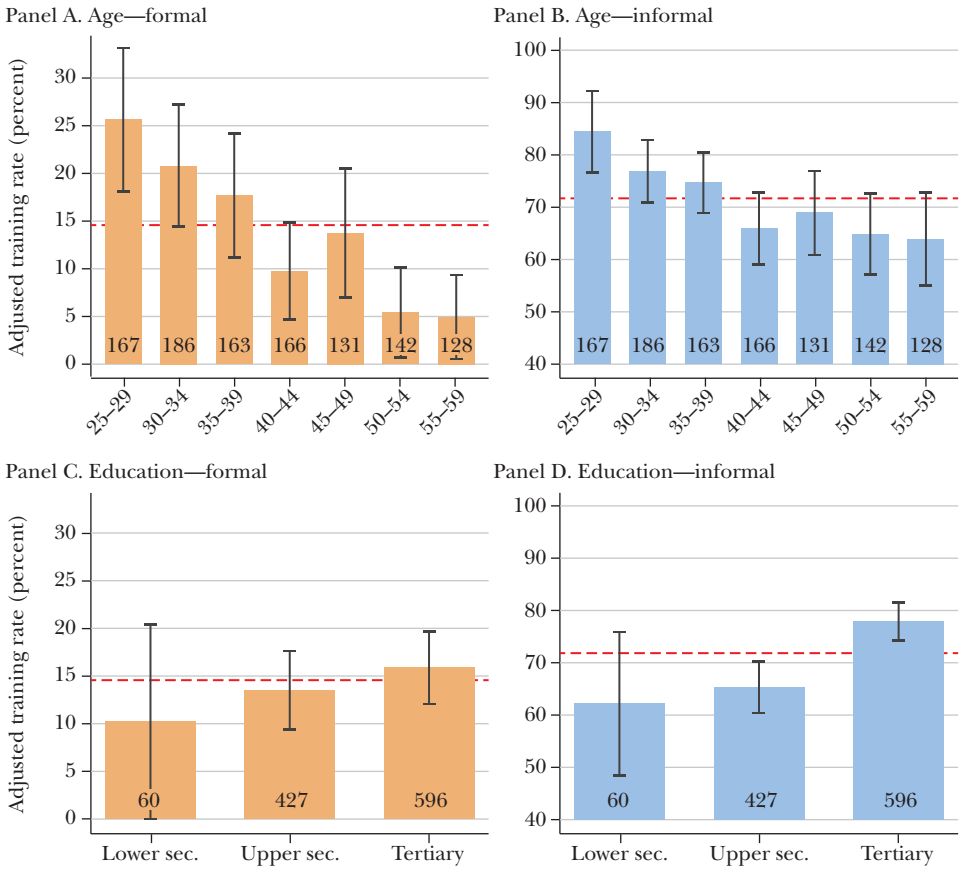
Programme for the International Assessment of Adult Competencies

The Programme for the International Assessment of Adult Competencies is a cross-national representative survey of adults administered by the Organization for Economic Co-operation and Development that measures adult skills and collects details on training participation, education, labor market status, job characteristics, and workplace practices. Specifically, we use data from cycle 2, fielded in 2022–2023. Our analysis follows the sample restrictions in Black, Skipper, and Smith (2023): ages 25–59, employees in the private or nonprofit sector, working full time (at least 30 hours per week). The US sample contains 1,083 respondents, of whom 1,068 have complete information on respondent and firm characteristics; the pooled cross-country sample varies between 15 countries (when using a full set of individual characteristics) and 29 countries (when using a smaller restricted set of variables that are available across the larger set of countries).

Figure 1 shows participation rates separately for formal training (study toward a recognized qualification) and informal training (any organized, nonformal training activity in the past twelve months, including open or distance courses, organized on-the-job training, workshops, and private lessons).[1] Across all age groups, the weighted participation rate over the last twelve months is 15 percent for formal training and 72 percent for informal training. Formal training is largely a subset of informal training: Of the 151 US workers in our sample (unweighted counts)

[1] We follow common usage in calling this *informal* training, though the Programme for the International Assessment of Adult Competencies labels it *nonformal*; the truly unstructured category—informal learning in PIAAC's sense—is not the focus of this section.

Figure 1
Correlates of Training: Age, Education, Establishment Size (US Data)
(adjusted training rates: formal (left) versus informal (right))

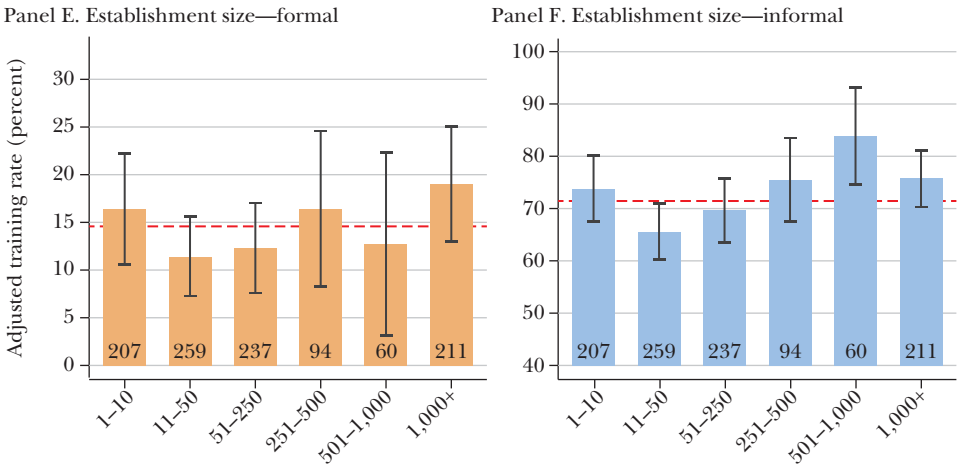


(continued)

who participated in formal training, 133 (88 percent) also participated in informal training, and only 18 did not. By contrast, informal training is often standalone: Of the 789 workers who participated in informal training, 656 (83 percent) did not also participate in formal training.

We follow Black et al. (2023) and examine the correlates of training participation with cumulative linear probability models, adding four blocks of variables sequentially: (1) individual characteristics (gender, age, education, dropout, marital status, native-born, number of children, child under six); (2) occupation (measured by the International Standard Classification of Occupations four categories) plus wage decile and manager status; (3) skill use at work (standardized Organization for Economic Co-operation and Development–derived indices for numeracy, reading, writing, and information and communications technology) plus an indicator for

Figure 1
Correlates of Training: Age, Education, Establishment Size (US Data)
(adjusted training rates: formal (left) versus informal (right)) (continued)



Source: PIAAC, cycle 2, 2022–2023.
Note: The figure plots adjusted training rates for formal and informal training by age, education, and establishment (workplace) size. The full specification controls for gender, age, education, dropout status, marital status, nativity, number of children, children under six, occupation dummies, manager status, wage decile, numeracy, literacy, writing, information and communication technology, teamwork, establishment size, part-of-larger-firm indicator, any workplace changes, industry dummies, learning at work, team voice, task discretion, and process improvement. All panels are estimated from the same full specification on the sample of respondents with complete information (United States; observations = 1,068), weighted by the Programme for the International Assessment of Adult Competencies final sample weight. Bars show 95 percent confidence intervals from heteroskedasticity-robust standard errors; design-based standard errors computed with PIAAC's 80 jackknife replicate weights are uniformly smaller, so the intervals shown are conservative. The dashed line is the overall weighted training rate. The numbers inside the bars are unweighted counts of respondents in each category in the full sample (observations = 1,083); establishment size is missing for 15 respondents, so the counts in the establishment-size panels sum to 1,068. Establishment size is the number of people at the respondent's work location.

team-based work; and (4) firm characteristics, including structural features (size, organizational reach, recent workplace change, industry fixed effects) and proxies for organizational practices experienced at work (standardized indices of learning at work, task discretion, team voice, and involvement in process improvement). The headline decomposition is estimated on the US sample.[2]

Both formal and informal training decline steeply with age. Workers aged 55–59 are about 21 percentage points less likely to participate in both formal

[2]Analogous models on the cross-country pooled sample appear in the Supplemental Appendix. The cumulative R^2 across the four blocks lets us ask how much of the variation in training participation traces to characteristics of the worker versus characteristics of the firm.

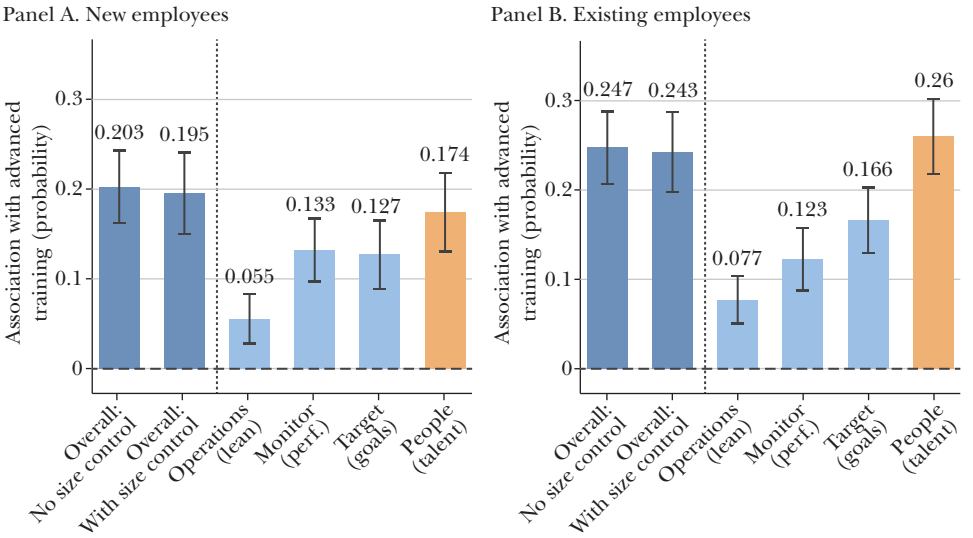
and informal training than workers aged 25–29, although the proportional decline is much larger for formal training. Relative to upper-secondary attainment (the reference category), tertiary educational attainment is associated with 13 percentage points higher informal training participation, but its correlation with formal training is modest and statistically insignificant once the full set of controls is included. Both gradients are broadly consistent with human capital theory. Younger workers have longer expected horizons over which they (and their employer) can recoup investments in skills. More educated workers may both demand and generate greater returns to additional skill acquisition, yet the education gradient also suggests complementarity, particularly for informal training: Workers who arrive with stronger foundational skills appear to receive more, not less, of the on-the-job skill investment that informal training represents. This pattern has implications for inequality, because firm-provided informal training would then amplify rather than equalize the human-capital differences workers bring with them when they enter the firm.

Firm characteristics add explanatory power on top of individual and job controls. Industry controls are jointly significant. Controlling for industry fixed effects, occupation, wage decile, manager status, and standardized Organization for Economic Co-operation and Development–derived skill-use indices (that seek to measure numeracy, reading, writing, and information technology skills), the individual establishment-size effects are jointly insignificant for formal training ($F = 1.17$, $p = 0.32$); informal training indicators, however, are jointly significant ($F = 2.72$, $p = 0.02$), driven by a single sharp differential at the 501–1,000 employee band. Focusing on establishment size (the measure available in the data) rather than firm size, however, masks potential economies of scale in training delivery. What predicts training far more consistently is whether the local establishment is part of a larger organization: +5.2 percentage points ($p = 0.04$) for formal training and +8.3 percentage points ($p = 0.005$) for informal training. This pattern is consistent with scale economies in training delivery: Larger organizations can amortize the fixed costs of curriculum design, deploy training infrastructure across sites, and operate the more complex production processes that raise the value of worker skills, in the spirit of the knowledge hierarchy argument we develop in the next main section. The evidence in Figure 1 also complements quasi-experimental evidence that workers who begin their careers at large firms enjoy persistent, long-run advantages by age 35 (Arellano-Bover 2024).

World Management Survey

The World Management Survey is fielded through double-blind, semistructured, open-ended telephone interviews with middle managers to measure the adoption of structured management practices across large samples of firms. Interviewers score firms on the adoption of 18 management practices related to operations, monitoring, target setting, and people (talent) management on a five-point scale, with higher scores denoting a higher and more systematic degree of adoption of

Figure 2
Management Practices and Advanced Training Provision (Coefficients from a Linear Probability Model)



Source: World Management Survey, 2022 wave.
Note: The figure plots linear probability model coefficients on advanced training for overall management score and four subcomponents (each entered separately). Details available in the Supplemental Appendix. The figure is split by training for new and existing employees. The sample is the World Management Survey 2022 manufacturing wave (observations = 1,344 firms per employee type). Coefficients are displayed in probability units (multiply by 100 for percentage points); bars show 95 percent confidence intervals from robust standard errors. All models control for country and two-digit industry fixed effects (with missing industry as its own category) and, where indicated, the logs of firm and plant employment with a missing-value indicator.

the specific practice (Bloom and Van Reenen 2010).[3] In 2022, the manufacturing survey was augmented with two questions asking 1,344 manufacturing firms across eleven countries about the type of training opportunities provided in the plant, distinguishing between no training, basic, and advanced (cross-functional) training to new and existing employees. Figure 2 examines how the provision of advanced training for new and existing employees varies with the management score, including controls for firm size and for industry and country fixed effects. The patterns help

[3] On monitoring, for example, a low-scoring firm tracks a handful of measures that do not reveal whether its objectives are being met and reviews them only among senior staff; a high-scoring firm tracks a defined set of indicators continuously, makes them visible to the people doing the work, and reviews them on a regular schedule. On target setting, a low score means that tracked goals are vague or purely financial, while high scores correspond to a balanced set of short- and long-run targets that individuals can connect to their own work. The individual scores are aggregated by simple averaging to obtain a single management score. Interviewers do not know the firm's performance and managers are not informed of the scoring scale.

to explain why, beyond establishment scale and organizational reach, some firms invest more in their workers than others.

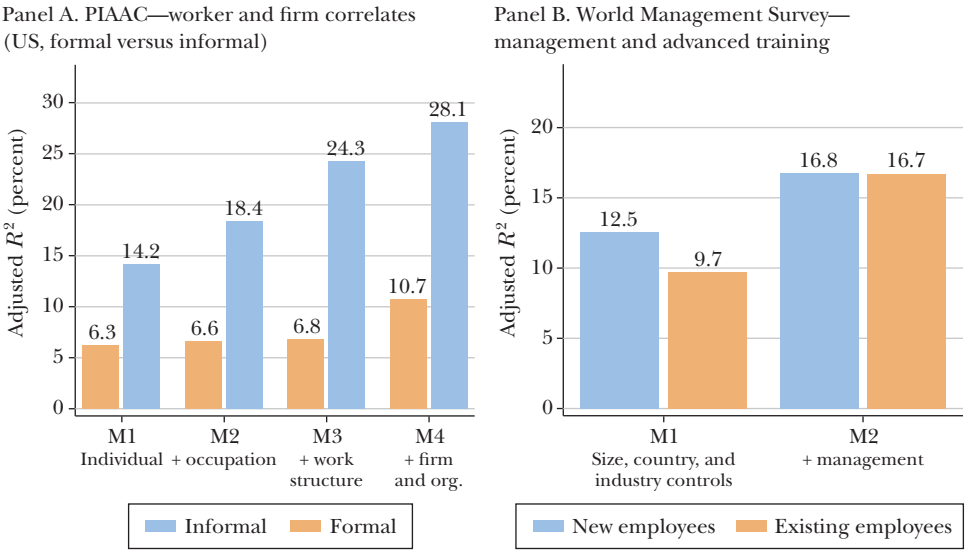
The key result is that firms with greater adoption of structured management practices (the first two bars in each panel) are significantly more likely than others to provide advanced training, even after controlling for firm size and industry and country fixed effects.[4] A one-point increase on the World Management Survey five-point overall management scale (roughly 1.7 standard deviations in this sample) is associated with an increase in the probability of providing advanced training of 19.5 percentage points (new employees) to 24.3 percentage points (existing employees). Decomposing the management index into its four subcomponents—operations, monitoring, targets, and people management—shows that people management, which captures the adoption of practices related to talent development, promotion, and performance evaluation, accounts for most of this association. In a specification that includes all four subcomponents jointly, a one-point increase in the people management score is associated with an 11-percentage point increase in advanced training for new employees and a 22-percentage point increase for existing employees, while the remaining subcomponents are weaker or statistically insignificant (monitoring retains a significant association for new employees). This pattern suggests that training is embedded in a broader system of people management processes. Firms that promote based on performance, evaluate talent rigorously, and address underperformance are also more likely to provide advanced training.

Figure 3 summarizes the explanatory power that firm-side variables add to standard worker and structural controls, using both the Programme for the International Assessment of Adult Competencies and World Management Survey datasets. Panel A shows the PIAAC decomposition: Adding the firm block (establishment size, organizational reach, industry fixed effects, and worker-reported organizational-practice proxies) on top of individual, occupation, and skill-use controls raises the adjusted R^2 from 6.8 to 10.7 percent for formal training (a 58 percent proportional gain) and from 24.3 to 28.1 percent for informal training (a 16 percent gain). The same qualitative ordering and approximate magnitudes appear in the great majority of Organization for Economic Co-operation and Development countries we examine.[5] Panel B shows the World Management Survey analog: Adding the overall management score on top of country, industry, and log-employment controls raises the adjusted R^2 for advanced training from 12.5 to 16.8 percent for new employees and from 9.7 to 16.7 percent for existing employees. Notably, management quality explains more of the training variation for *existing* employees than for *new* employees,

[4] Size and productivity are positively correlated in the cross section, but dynamic firms may have better management practices in place to prepare for growth.

[5] In a specification covering 29 countries (which omits the Organization for Economic Co-operation and Development–derived skill-use indices and other individual variables that are not included uniformly in public-use files), the average gain in adjusted R^2 from adding firm and organizational-practice variables is roughly 11 percentage points for informal training—nearly an order of magnitude larger than what individual characteristics alone explain. Korea is the lone outlier, with very low explanatory power across all blocks. Details are available in the Supplemental Appendix.

Figure 3
Explaining Training Participation and Provision



Source: PIAAC and World Management Survey
Note: The figure plots adjusted R^2 from cumulative four-block decompositions for the Programme for the International Assessment of Adult Competencies sample in panel A (Linear Probability Models, US sample) and the World Management Survey incremental contribution to adjusted R^2 from adding management practices in panel B. All models within a panel are estimated on a constant sample (PIAAC United States, observations = 1,068; World Management Survey, observations = 1,344 firms per employee type); panel B adds the overall management score to country, industry, and log-employment controls. Calculations are available in the Supplemental Appendix.

consistent with the view that ongoing worker development requires deeper organizational capability.

Although the two panels use different operationalizations of firm characteristics, they tell a coherent story: Firm-level variation—and management quality in particular—adds meaningful explanatory power, even though most of the variation in training participation remains unexplained by observables in either dataset.

Where a person works and how their firm is organized are among the strongest predictors of whether they receive training. Although this evidence is only descriptive, it is consistent with quasi-experimental work that growth of wages and human capital depends to some extent on the firm in which a person ends up working. Recent examples include Arellano-Bover (2024), who uses labor-market entry conditions as a source of exogenous variation to identify firm effects on young workers' careers, and Arellano-Bover and Saltiel (2026), who use worker-mover designs (that is, shifts in wage growth as workers move, or do not move, between firms) to identify firm-specific human-capital growth. A growing body of evidence establishes that firm heterogeneity is quantitatively important in explaining adult

earnings inequality.[6] Our results suggest that firms' investments in their workers' skills, and the way they organize those investments, are at least partially responsible.

Employer Reskilling and Upskilling Survey

To look more closely at what training bundles contain, and to examine how firms organize training itself, we turn to a 2023 survey of chief human resources officers and senior human resources leaders at 954 US firms with 100 or more employees (Doûmi, Sadun, and Tamayo 2026). The survey was administered in July 2023 by Dynata, a professional survey provider, to one senior human resources leader per firm, with quotas on firm size, industry, and geography. The survey asks each respondent about a single flagship reskilling or upskilling program in detail (its participants, delivery model, requirements, funding, and incentives provided to trainees and their managers) together with the firm's broader human resources infrastructure. The 954 firms we analyze are those that report operating at least one upskilling or reskilling program.

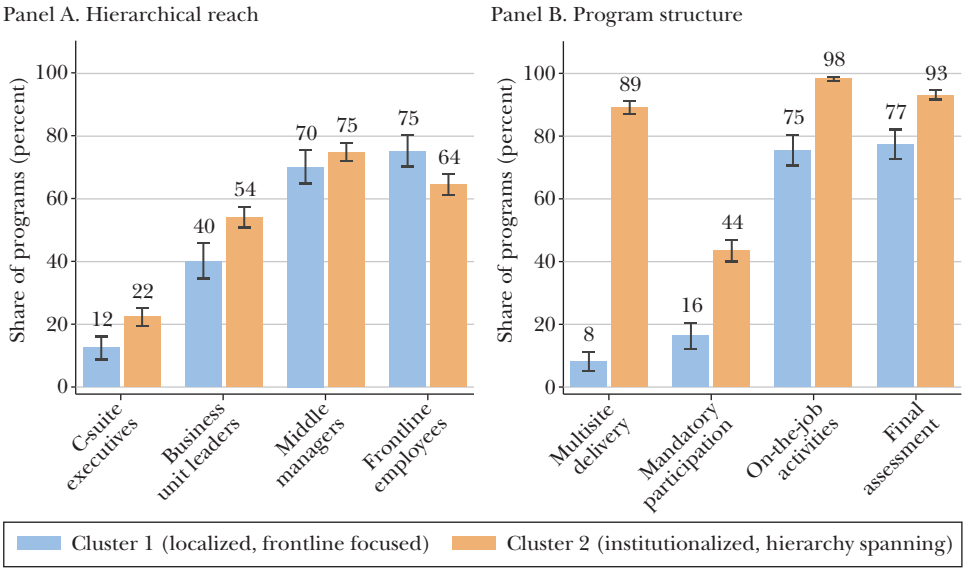
Compared with samples used in the Programme for the International Assessment of Adult Competencies and World Management Survey, the strength of these data is their granularity; they let us look beyond *whether* firms train and ask *how* they organize that investment. The trade-off is that the sample skews toward larger firms with some training in place. Hence, the data are best positioned to understand differences in how programs are targeted at different workers or designed and administered, rather than whether a firm trains at all.

We group firms into two clusters based on the characteristics of their flagship program, as shown in Figure 4. The groups come from a cluster analysis in which each firm is described by 86 program design features covering delivery, participant selection, funding, incentives, assessment, and placement.[7] The first cluster (about one-quarter of firms) runs programs that are local, in scope, and oriented toward frontline employees. They tend to take place at a single site, are typically voluntary, are most often funded entirely by the organization, and are disproportionately oriented toward rank-and-file workers. The second cluster (about three-quarters of firms) runs programs that are much more structured in delivery. Nine in ten of these programs run across multiple locations, relative to fewer than one in ten of the first cluster. Nearly all include on-the-job activities (98 against 75 percent), most culminate in a final assessment or certificate (93 against 77 percent), and over 40 percent are mandatory rather than voluntary (44 against 16 percent). They reach farther up the hierarchy as well: 22 percent explicitly include C-suite executives, and 54 percent include business unit leaders, against 12 and 40 percent in the first cluster.

[6] Estimates of this kind should be read with care. The firm effects recovered from two-way fixed-effects models in the tradition of Abowd, Kramarz, and Margolis (1999) also absorb sorting and match quality, so they are not a clean measure of a firm premium (Borovičková and Shimer 2024).

[7] Because those features mix binary and continuous items, dissimilarity is measured with the Gower metric, and firms are grouped by agglomerative hierarchical clustering; standard criteria for choosing the number of groups favor a two-group split. The Supplemental Appendix reports robustness checks, which suggests the two configurations are best read as poles of a continuum rather than sharply bounded types.

Figure 4
Firms Cluster in How They Deliver Training



Source: Employer Reskilling and Upskilling Survey.
Note: Survey of chief human resources officers and senior human resources leaders, fielded mid-2023, at US firms with 100 or more employees. Observations = 954 firms with structured upskilling or reskilling programs (the cluster sample). Firms whose only training is general or job induction are excluded. Cluster 1 contains 260 firms and cluster 2 contains 694; three cluster 1 firms did not answer the specific items shown in the figure (the Gower-based clustering tolerates item-level missingness), so the bars use 257 and 694 responses. Bars show the share of firms in each cluster whose program has the listed feature, with 95 percent confidence intervals.

A core finding from this survey is that the organization of training is closely tied to a firm's formal human-capital infrastructure. Skill taxonomies, skill inventories, and internal learning academies are the strongest predictors of whether firms have more structured investments in training. Skill taxonomies and skill inventories are associated with 27 and 17 percentage point increases, respectively, in the probability that a firm runs the more structured type of program (that is, of belonging to the second cluster).[8] Multinationals are 13 percentage points more likely than domestic firms to run such programs. Firms with 10,000 or more employees are 14 percentage points more likely than firms with 100 to 999 employees to have structured training programs. Formal human resources infrastructure, multinational scope, and scale

[8] In the Supplemental Appendix, we estimate linear probability models of membership in the second (more structured) cluster that hold constant firm size, multinational status, public listing, union density, the firm's stated rationale and goals for investing in training (upskilling or reskilling, and the objectives cited), its skill-gap profile, its broader human-resources infrastructure, industry dummies, and respondent characteristics.

together account for most of the explanatory variation; among them, the human resources practices are the most important.

Overall, the firms running structured, hierarchy-spanning programs are precisely the firms with the organizational capacity that can make such investments worthwhile. Conversely, the firms running localized, frontline-focused programs are those with less formal human resources infrastructure and narrower organizational scope. The organization of training, in other words, correlates with the broader organizational design of the firm.

Summary

The figures in this section show four patterns that motivate our discussion in the next section of the organizational economics of training. First, firm-provided training is increasing in workers' education, suggesting a complementarity between on-the-job learning and formal education and skilling. Second, firm characteristics add substantial explanatory power beyond worker observables, and this pattern generalizes across countries, labor market institutions, and industries. Third, better-managed firms, especially those with stronger people-management practices, are closely linked to advanced training provision. Fourth, firms differ not only in *whether* they train but also in *how* they organize training. Training appears as part of a broader portfolio of organizational design choices, in which formal human resources infrastructure, hierarchical reach, and program institutionalization move together.

Our results suggest that firms differ in management capability and organizational design, and that these differences shape the returns to skill within the hierarchy. Heterogeneity in training investment is, therefore, not merely a response to worker demand or labor market frictions, but an endogenous component of how firms produce and compete. In that view, adult skill formation is tightly related to both individuals accumulating human capital, and firms deciding when, how, and for whom additional skill is valuable.

In the next section, we sketch a framework to help understand what drives firms facing the same conditions to make different choices about how to organize. The literature on the determinants of management practices has traditionally focused on factors like labor market conditions or competition, but other factors emphasized in some papers, like family ownership, suggest that the identity of a top manager also tends to matter (Bloom, Sadun, and Van Reenen 2010). We argue that differences in top-manager skill can explain differences in organizational structure and people-management practices and, through these, the heterogeneity in training investment we have just documented.

Training as an Organizational Design Choice

Standard theories of firm training grapple with the question of why firms might perceive such training as in their interest, rather than just raising the wages

of workers who may leave for an alternative employer. Becker (1964) famously argued that firms would only pay for firm-specific skill acquisition, not general training. Later theories emphasized labor market frictions. Monopsony and wage-compression models such as in Acemoglu and Pischke (1999) predict more training when firms have greater wage-setting power. When search frictions or labor market power compress the wage schedule, a worker's wage rises by less than their productivity, enabling firms to invest in skills and capture value. However, these models do not explain why firms from the same industries and local labor markets would make different training investments. The skill-weights framework developed by Lazear (2009) allows firms to value different bundles of general skills, making those bundles of skills effectively firm-specific, but takes skill weights for a firm as given rather than explaining their origin. Models of asymmetric information about worker productivity explain how firms capture rents on training investment that are unobserved by outside employers, but also do not explain heterogeneous firm investments.

More fundamentally, we argue that this framing of the firm's decision to offer training elides a basic point. Firm training should be viewed as an endogenous design choice. Organizational differences necessitate different training investments, which in turn depend on managerial capabilities. Firms that do more training prepare workers for organizational complexity; firms that do less training also do less customization, enabling the training firms to capture value.

Knowledge Hierarchies

We build on a long tradition of models in organizational economics, particularly the "knowledge hierarchy" model of Garicano (2000) and the extensions by Caliendo and Rossi-Hansberg (2012) and others. To fix ideas, consider the knowledge hierarchy of a management consulting firm. When a client engagement begins, the day-to-day analytical work lands first with junior consultants. They conduct interviews, build spreadsheets, and structure analysis. Most problems that arise are well within their competence, but when junior consultants encounter an issue that exceeds their skill, they ask a senior consultant. If the problem is harder still, the senior consultant raises the question with the partner who owns the relationship. In this knowledge hierarchy, the partner handles only the most difficult problems. The defining feature of a knowledge hierarchy is that problems are indexed by difficulty, and only the higher levels of a pyramidal organization can solve the more difficult problems.

A knowledge hierarchy is an efficient way to deploy the partner's time and judgment, which is the firm's scarcest labor resource. The upward filtering of harder problems that happens via the knowledge hierarchy enables managerial leverage. A senior consultant who can resolve 70 percent of the problems that arrive passes only 30 percent upward; one who can resolve 60 percent passes 40 percent, increasing the burden on the next layer. Worker training that improves autonomous problem resolution creates efficiencies in upper layers, because fewer problems arrive there. This insight is not unique to consulting and other professional services. Technology firms, government agencies, and manufacturing firms (which Caliendo, Monte,

and Rossi-Hansberg (2015) study through the lens of hierarchies) are organized in layers for many of the same reasons. The conditions hold whenever exceptions are escalated, like when a nurse asks an attending physician for help on a complicated patient or when a loan officer refers an anomalous balance sheet to a supervisor. The value of the time savings from training a lower-level worker simply depends on the cost of the upper layer, which in turn depends on the skill of the people there and, ultimately, on the skill of whoever sits at the top.

Empirical work is consistent with the importance of this kind of filtering mechanism. For example, when a Colombian government organization trained lower-level workers, the volume of help requests filtering up through the hierarchy fell, allowing managers to shift their time from supporting existing workers to other tasks (Espinosa and Stanton 2026). However, we have not yet explained why managers with different abilities make different organizational choices.

How Manager Ability Shapes the Firm

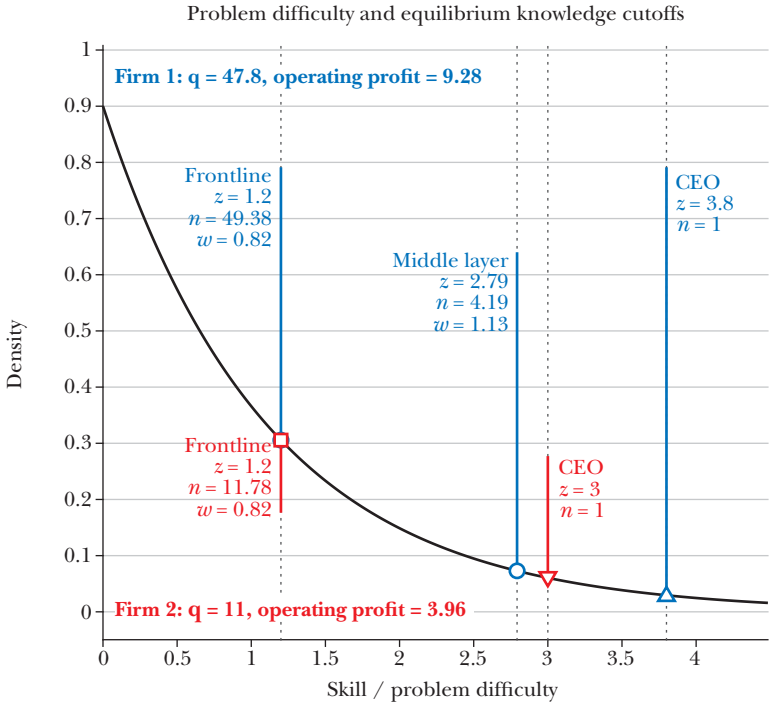
To explore why managers build different hierarchies, we examine how top managers with different abilities organize their firms. We make one key change to the canonical knowledge hierarchies framework, which we lay out formally in Deming, Gilman, Sadun, and Stanton (2026). Our illustration builds on Garicano and Rossi-Hansberg (2006), who show that the most skilled workers in the economy will start firms when costs of knowledge acquisition vary between individuals. To build intuition, we focus on a stylized industry equilibrium where two entrepreneurs with the highest skills in the economy start firms. We set these workers' skills at a high level, while assuming that a fixed cost of entry prevents a third entrepreneur with lower skills from forming a third, organized firm. However, workers can always produce as individuals, selling their own solutions. Other than the candidate entrepreneurs, everyone else in the economy starts with identical skills, which will potentially give rise to either firms or workers paying for training.

The key difference in this model from the literature involves wage determination. In our model, workers' pay is based on their outside option. Firms have a training incentive if melding a worker's skill to fit the organization's needs has a higher payoff than if the worker went to a rival or if the worker produced alone as a sole proprietor.[9] Firms with simpler organizations are less likely to have an idiosyncratic match-specific fit that beats a rival firm and a worker's option to produce independently, explaining why competing firms make different training choices.

For concreteness, consider two consulting firms operating in the same city, competing for the same clients and to hire the same staff. The more capable partner—who we will call the founder of firm 1—can solve a wider range of

[9] This follows a tradition in labor economics from Postel-Vinay and Robin (2002) that treats the wage determination process like an auction. As far as we know, papers in the knowledge hierarchies literature have used different wage determination processes. Our treatment differs from the model in Postel-Vinay and Robin (2002), which features a competence parameter on the worker side, as well as recent contributions on worker and firm dynamics by Gualdani, Pastorino, de Paula, and Salgado (2026) and Di Addario, Kline, Saggio, and Sølvsten (2023), by focusing on firms' investment incentives in light of the wage-setting mechanism.

Figure 5
Optimal Hierarchy Structure for Two Firms with Different Entrepreneur Skill Levels



Source: Authors' calculations.
Note: The figure plots equilibrium organizational structure for firm 1 and firm 2 against the density of problems in the economy. Skills are represented as segments of the real line and workers with skill z can solve any problem that has a difficulty index less than their skill. Workers are born with skill equal to 1.2, such that schooling prepares workers with exactly what is required to be a sole proprietor, while further training costs to acquire new skills are proportional to the interval length of the change in skill. There is a mass of 100 workers that can be hired by firms or work alone. Firms and individuals compete in quantities, taking into account a linear, downward sloping industry demand curve. Labels in blue show the skill levels (z), number of workers (n), and wages (w) for workers at different layers for firm 1. Labels in red show the same thing for firm 2. Firm 1's quantity is 47.8 and operating profits net of training costs are 9.28. Firm 2's quantity is 11 and operating profits are 3.96.

problems than the founder of firm 2. However, if the founder of firm 1 is constantly interrupted by problems that a strong senior consultant could resolve, the firm is wasting its advantage. The firm with the more capable manager has more to gain from building a strong middle layer, so that fewer problems filter upward. Similarly, a more capable manager can scale personal influence with a larger frontline workforce, but only if the intermediate layers beneath can handle the increased volume. Firms with better top managers have a greater incentive to build out structure, add middle layers, and adopt the talent management practices that empirically correlate with deep training investments.

The key result is a complementarity between managerial talent and organizational depth: The more capable the leader sitting atop the firm, the more layers it builds. Figure 5 considers a benchmark numerical example to illustrate the relationship between the number of layers in an organization, top manager talent, and training. We consider a mass of 100 workers that can be allocated between the firms or solo operations, but we have stripped away all product-side differentiation. Firms or sole proprietors compete based simply on the number of solutions to problems they can solve, and we look for a Cournot equilibrium in quantities where firms choose how to organize (worker skills, numbers, and layers) and sole proprietors make their own skill investments. Frontline workers and sole proprietors address problems that vary in difficulty, illustrated by the density curve in Figure 5, which we parameterize as draws from an exponential distribution with parameter $\lambda = 0.9$. The respective CEO skills, 3.8 and 3.0, imply that the leader of firm 1 can solve $1 - e^{(-0.9 \times 3.8)} = 97$ percent of problems, while the leader of firm 2 can solve 93 percent. Sole proprietors and frontline workers have skills of 1.2, enabling them to solve 66 percent of problems they draw.

The relatively greater ability to solve exceptional problems leads to very different organizations. Firm 1's more capable chief executive officer operates at a much larger scale, with about 49 frontline workers, compared to firm 2, which hires only 12 frontline workers. The more capable entrepreneur (firm 1) builds a deeper hierarchy with a middle layer of managers above frontline workers, while the less capable entrepreneur (firm 2) operates a flatter structure with only frontline workers under the entrepreneur. By training a relatively small middle layer (here 4.2 workers with skill 2.8) to filter problems beyond frontline workers' capacity, firm 1 can support more frontline workers while the time constraint of the chief executive officer just binds due to problems that require ability beyond the middle layer's skills. Ultimately, firm 1 produces a quantity of 47.8 problems that is more than four times as large as firm 2 because of the complementarity between organization and managerial skill. Each of these choices is stable for the firms. Adding a fourth layer to firm 1 and adjusting the skill mix and numbers across layers would add more payroll and training expenses, reducing profit to 8.15 from 9.28 with three layers. If firm 2 were to add a third layer and reoptimize, profit would fall from 3.96 to 3.77.

This framework illustrates the endogenous relationship between managerial capability and organization design. A strong senior consultant is worth more as a buffer beneath a highly capable partner than beneath a less capable one, simply because the partner's time is more valuable. This result holds with a simple conception of "skill" that is generally valued in the labor market: the worker's cumulative knowledge for handling consulting tasks of varying degrees of complexity.

Wages, Training Incentives, and the Optimal Training Investment

Before turning to the training decision itself, we need to ask whether firms have any incentive to train in the first place or whether trained workers will be poached by other firms or, alternatively, leave to set up their own shops.

The allocation in Figure 5 is stable because the complementarity between worker skill and organizational structure makes a trained worker more valuable inside firm 1 than in any other firm or working on their own. Firm 2 might want to poach the trained employee from firm 1, but it is only willing to pay for the skill it can actually use. However, the training in firm 1 was calibrated to the needs of a hierarchy built beneath a skilled partner, and those needs are not replicated in firm 2's shallower structure.[10] The trained middle layer workers' wages of 1.13 reflect what they would make as a sole proprietor with these skills, but they do earn a premium relative to the wage of 0.82 that they would make without the firm's training investment. However, the worker is more valuable inside firm 1 than outside, because middle-level workers can handle more problems than if they had to work a problem end to end. Firm 1 therefore earns rents on its training investment—not because the labor market is imperfect in any standard sense, but because the organizational context itself creates value that is not portable. The labor market is effectively specific, even though the skills are not. In this example, skills are entirely one dimensional and general; the weights on skills are not combinations of different skills, but instead arise because some firms are willing to pay to make or buy a more capable generalist.

Given that the firm has an incentive to train, how does it decide where to invest? Consider firm 1's decision about whether to train its senior consultants—the middle layer of its hierarchy. Every senior consultant who receives training resolves more of the problems escalated to the partner by junior consultants, so fewer problems reach the partner's desk and the partner can handle more of them, raising firm productivity. The firm trains when the gains from protecting the partner's time exceed the cost of raising the skills of every person in that layer.

The same logic applies one layer down, but the trade-offs change. Training the entire group of frontline workers pays off in a smaller layer above—the savings accrue to the senior layer rather than directly to the partner. Whether this is worthwhile for the firm depends on how expensive the senior layer is and how many juniors each senior oversees. In firm 1, where the senior layer is more skilled and the ratios steeper, the firm still concentrates its investment in the upper layers where the bottleneck is most binding. This insight mirrors the clustering of training strategies in the survey of chief human resources officers: Some firms train mainly managers, others focus on frontline workers, and the pattern reflects organizational structure rather than arbitrary preference.

Two further implications follow. First, firms with more capable leadership invest more in training overall and capture larger rents from doing so—not because

[10] The reason a trained worker can be worth more inside the firm than as an independent provider is due to leverage: A trained worker can support several workers in the hierarchy below, which can be more valuable than individually solving a problem. In knowledge hierarchies models, leverage is determined by a "helping cost" parameter, which represents the fraction of a time period that a higher-layer worker spends addressing a problem that gets passed up. Lower-layer workers and sole proprietors spend their time on a single problem. In the numerical example we discuss, we set the helping cost parameter, h, to 0.25, following Garicano (2000).

workers are stuck, but because the trained worker is genuinely worth more inside a hierarchy with more capable leadership. Firm 1's trained middle-layer workers earn a higher wage (1.13 rather than 0.82) than they would have earned without training; this wage exceeds what firm 2 would pay for their skills and equals their outside option as sole proprietors with those skills. The organizational complementarity benefits both sides because these workers' skills expand after firm investment. Second, a survey that records only whether a firm has a training program will miss the most important variation: how much it invests, in which layers, and why.

What the Framework Explains and What It Misses

In this framework, differences in firm training investments are driven by variation in managerial capability and endogenous organizational design choices. The framework also explains the long-standing empirical puzzle that firms often pay for their workers to receive general skills training and that these workers generally stay at the firm that trained them. In addition, there is a sharp prediction that training investments often matter most in the middle and upper layers of the management hierarchy.

In standard monopsony or search-friction models, firm rents from training arise because the market works imperfectly—workers cannot readily or at low cost find alternative employers that would pay them their true value. In contrast, in this hierarchy model, rents arise because the organizational context itself creates value that is not fully portable. The thinness of the posttraining labor market is endogenous to the organizational structure. This distinction has policy implications. If training incentives arise from market imperfections, then competition policy and labor market policies that encourage training may be at odds. On the other hand, if training arises from organizational heterogeneity, public policies to promote worker training that ignore firm heterogeneity and organizational complementarities may see very low returns from generic training investments.

More broadly, firms that invest more in training also create richer internal labor markets, with more rungs on the promotion ladder and faster potential wage growth for workers who enter the firm. This finding is consistent with natural experiments where early career sorting into high-wage-growth firms generates persistent wage differences (for example, Arellano-Bover and Saltiel 2026). Better-managed firms probably also invest more in leadership development. The people-management practices captured in the World Management Survey include monetary and nonmonetary rewards for high performance and opportunities for internal promotion, both of which reflect leadership development investments.

What happens to firm training if artificial intelligence alters incentives to hire and train junior workers? Ide and Talamàs (2025) develop a knowledge hierarchies model where AI agents can act either as copilots or autonomous coworkers. AI that acts as a copilot reduces inequality, while autonomous AI increases it. AI might also blunt incentives for training by substituting for the rote work that formerly paid for the training of junior workers. In the model of Garicano and Rayo (2026), the effect of AI depends on how much it raises the minimum skill at which an

entry-level worker becomes useful (the "entry floor") versus increasing the value of a fully trained worker (the "expert ceiling"). In our framework, the impact of AI also depends on the organizational surplus created by embedding a trained worker beneath scarce managerial talent.

This distinction generates firm heterogeneity in the response to increasing AI capabilities. Firms with simple organizations will see the training period compress and eventually disappear, as in the model of Garicano and Rayo (2026). If training takes time and exposure, rather than the instantaneous investment our static framework considers, firms with talented managers and strong organizational complementarities may continue to train even after junior production ceases to finance the investment. However, training in these firms may change, with fewer entrants, a higher "entry floor," more intensive and formal instruction, and greater concentration on the middle and upper layers of the hierarchy.

One limitation of our framework is that it treats these internal labor markets as arising from managerial quality differences alone, but there may also be economies of scale and other sources of comparative advantage in training provision that shape which firms invest in workforce development. Some organizations may simply be better than others at designing, organizing, and delivering training programs at low cost. If on-the-job experience and formal learning are complements, then firms may need to participate actively in skill formation even when the underlying skills are portable across employers. In that environment, firms with superior organizational capabilities in training provision may offer development opportunities as a workplace amenity that helps attract and retain workers (Maestas et al. 2023).

Our framework is not intended to provide a comprehensive explanation for all forms of firm-sponsored training. Instead, it is designed primarily to illustrate "what is in it for firms" when they invest in worker skill. Other mechanisms may support employer-sponsored learning, even absent strong organizational complementarities. What is important is the interaction between internal structure and outside wages or opportunities, a logic that also underlies why firms train more when they have noncompete agreements at their disposal (Starr 2019) or when contracts with repayment provisions are available (Hoffman and Burks 2020). These arrangements effectively allow firms to finance general skill acquisition while partially protecting themselves against immediate worker turnover.

Finally, the basic hierarchy framework intentionally abstracts from worker heterogeneity, treating workers as homogeneous before training and workplace learning occur. Relaxing that assumption would likely generate sorting and matching patterns that interact importantly with firms' training incentives. We conjecture that models with heterogeneous workers and heterogeneous firms would generate positive assortative matching between highly educated or highly skilled workers and firms with the greatest skill demands and strongest incentives to invest in training. In such environments, firms with the highest managerial capability or organizational complexity may both select and develop workers with the greatest learning potential, amplifying observed differences in career trajectories across firms and helping to explain the positive relationship between education and on-the-job training.

Conclusion

Firm-sponsored general skills training is very common. Much of the literature has explained this phenomenon through frictional labor markets. For example, in studying how Indian garment firms can support the provision of social skills training, Adhvaryu et al. (2023) find extremely high returns on investment. They point to labor market imperfections to explain the minimal effects on wages or retention. However, the garment sector in India has many firms, raising a puzzle about how training can operate in sectors with many firms, each of which makes different training investments.

In this paper, we have suggested an alternative form of labor market friction that stems from organizational design. In the social skills context, the gains from training appear to come from improved collaboration and team effectiveness, alleviating the need for managerial attention. If all firms had the same organization, we would expect these sources of advantage to dissipate. Instead, heterogeneous organization generates different skill demands from workers—whether in the Indian garment industry or elsewhere—which make general skills effectively firm-specific and justify firms' investment in training.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67 (2): 251–333.

Acemoglu, Daron, and Jörn-Steffen Pischke. 1999. "The Structure of Wages and Investment in General Training." *Journal of Political Economy* 107 (3): 539–72.

Adda, Jérôme, and Christian Dustmann. 2023. "Sources of Wage Growth." *Journal of Political Economy* 131 (2): 456–503.

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2023. "Returns to On-the-Job Soft Skills Training." *Journal of Political Economy* 131 (8): 2165–208.

Arellano-Bover, Jaime. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size." *Journal of Labor Economics* 42 (2): 549–89.

Arellano-Bover, Jaime, and Fernando Saltiel. 2026. "Differences in On-the-Job Learning across Firms." *Journal of Labor Economics* 44 (1): 149–88.

Barron, John M., Mark C. Berger, and Dan A. Black. 1999. "Do Workers Pay for On-the-Job Training?" *Journal of Human Resources* 34 (2): 235–52.

Becker, Gary S. 1964. *Human Capital: A Theoretical Empirical Analysis with Special Reference to Education.* University of Chicago Press.

Black, Dan A., Lars Skipper, and Jeffrey A. Smith. 2023. "Firm Training." In *Handbook of the Economics of Education*, Vol. 7, 287–468. Elsevier.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2010. "Recent Advances in the Empirics of Organizational Economics." *Annual Review of Economics* 2: 105–37.

Bloom, Nicholas, and John Van Reenen. 2010. "Why Do Management Practices Differ across Firms and Countries?" *Journal of Economic Perspectives* 24 (1): 203–24.

Bloom, Nicholas, Renata Lemos, Daniela Scur, Rafaella Sadun, and John Van Reenen. 2022. "World Management Survey, Manufacturing (2022 wave) [database]." World Management Survey. https://worldmanagementsurvey.org.

Borovičková, Katarína, and Robert Shimer. 2024. "Assortative Matching and Wages: The Role of Selection." NBER Working Paper 33184.

Brynjolfsson, Erik, Bharat Chandar, and Ruyu Chen. 2025. "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence." Unpublished.

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg. 2015. "The Anatomy of French Production Hierarchies." *Journal of Political Economy* 123 (4): 809–52.

Caliendo, Lorenzo, and Esteban Rossi-Hansberg. 2012. "The Impact of Trade on Organization and Productivity." *Quarterly Journal of Economics* 127 (3): 1393–467.

Deming, David J. 2025. "Why Do Wages Grow Faster for Educated Workers?" *Journal of Labor Economics*. https://doi.org/10.1086/737990.

Deming, David J., Julia Gilman, Raffaella Sadun, and Christopher Stanton. 2026. "Organization as a Rationale for the Provision of General Skills Training." Unpublished.

Deming, David J., Rafaella Sadun, and Christopher Stanton. 2026. *Data and Code for "The Role of Organization in Firm-Sponsored Training."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E253655V1.

Di Addario, Sabrina, Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten. 2023. "It Ain't Where You're From, It's Where You're At: Hiring Origins, Firm Heterogeneity, and Wages." *Journal of Econometrics* 233 (2): 340–74.

Doûmi, Leila, Rafaella Sadun, and Jorge Tamayo. 2023. "Employer Reskilling and Upskilling Survey (ERUS)." Harvard Business School. Restricted-access data.

Doûmi, Leila, Raffaella Sadun, and Jorge Tamayo. 2026."How Do Firms Train Their Workforce in the Age of Technology? Evidence from a Large-Scale Survey of U.S. Firms." Unpublished.

Espinosa, Miguel, and Christopher Stanton. 2026. "Training, Communications Patterns, and Spillovers inside Organizations." *Journal of Political Economy* 134 (7): 1933–66.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108 (5): 874–904.

Garicano, Luis, and Luis Rayo. 2026. "Training in the Age of AI: A Theory of Career Viability." CEPR Discussion Paper 20634.

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics* 121 (4): 1383–435.

Gibbons, Robert, and Michael Waldman. 2006. "Enriching a Theory of Wage and Promotion Dynamics inside Firms." *Journal of Labor Economics* 24 (1): 59–107.

Gualdani, Cristina, Elena Pastorino, Áureo de Paula, and Sergio Salgado. 2026. "Identifying Uncertainty, Learning about Productivity, and Human Capital Acquisition: A Reassessment of Labor Market Sorting and Firm Monopsony Power." NBER Working Paper 34973.

Hoffman, Mitchell, and Stephen V. Burks. 2020. "Worker Overconfidence: Field Evidence and Implications for Employee Turnover and Firm Profits." *Quantitative Economics* 11 (1): 315–48.

Hosseini Maasoum, Seyed Mahdi, and Guy Lichtinger. 2025. "Generative AI as Seniority-Biased Technological Change: Evidence from US Résumé and Job Posting Data." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.5425555.

Ide, Enrique, and Eduard Talamàs. 2025. "Artificial Intelligence in the Knowledge Economy." *Journal of Political Economy* 133 (12): 3762–800.

Lazear, Edward P. 2009. "Firm-Specific Human Capital: A Skill-Weights Approach." *Journal of Political Economy* 117 (5): 914–40.

Maestas, Nicole, Kathleen J. Mullen, David Powell, Till von Wachter, and Jeffrey B. Wenger. 2023. "The Value of Working Conditions in the United States and Implications for the Structure of Wages." *American Economic Review* 113 (7): 2007–47.

OECD. 2023. "PIAAC 2nd Cycle Database: Public Use Files and International Codebook." Paris: OECD. https://www.oecd.org/en/data/datasets/PIAAC-2nd-Cycle-Database.html (accessed February 2026).

Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica* 70 (6): 2295–350.

Sandvik, Jason, Richard Saouma, Nathan Seegert, and Christopher Stanton. 2026. "Should Human Capital Development Programs Be Mandatory or Voluntary? Evidence from a Field Experiment on Mentorship." *Management Science* 72 (6): 4610–32.

Starr, Evan. 2019. "Consider This: Training, Wages, and the Enforceability of Covenants Not to Compete." *ILR Review* 72 (4): 783–817.

Journal of Economic Perspectives—Volume 40, Number 4—Fall 2026—Pages 207–214

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 e-mail at <taylort@macalester.edu>, or c/o Journal of Economic Perspectives, Macalester College, 1600 Grand Ave., Saint Paul, MN 55105.

Smorgasbord

Tobias Adrian offers a morphology of "Tokenized Finance" (International Monetary Fund, Note/2026/001, April 2026, https://www.imf.org/-/media/files/publications/imf-notes/2026/english/insea2026001.pdf). "Financial sector policy frameworks have evolved around institutions, balance sheets, and markets that operate with temporal frictions: end-of-day settlement, batch processing, and delayed reconciliation. These frictions are not only costly to end-investors, but they also provide temporal buffers that allow exposures to be netted, liquidity to be mobilized, and authorities to intervene before settlement becomes final. . . . Tokenization enables financial claims—including money, securities, and derivatives—to be represented as programmable digital tokens recorded on shared ledgers. This capability allows for real-time atomic settlement, collapsing multiple stages of the traditional

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20261531.

financial value chain into a synchronized process . . . Tokenization challenges crisis management and resolution frameworks that are built around nationally domiciled institutions, territorially bounded infrastructures, and jurisdiction-specific legal authority. In tokenized systems, transactions are executed on shared ledgers spanning multiple jurisdictions, allowing assets, liabilities, and collateral to move across borders at machine speed and without a clear geographic anchor. This creates a fundamental mismatch between the global, continuous operation of tokenized finance and resolution regimes that rely on jurisdictional control over institutions and locally situated assets, as the key levers of control may instead lie in governance keys, consensus mechanisms, or smart contract logic operating across borders. . . . Legal uncertainty is a major barrier to scaling tokenized systems beyond pilot projects. Market participants require clarity on whether tokenized records constitute a definitive proof of ownership and whether the settlement finality achieved on a ledger is legally recognized. Without such clarity, tokenized markets risk remaining fragmented and peripheral. A dual-layer legal approach is likely to emerge as best practice. Smart contracts define operational rules, whereas traditional legal agreements establish rights, obligations, and dispute resolution mechanisms. Legislators and courts must clarify the relationship between these layers, ensuring that legal certainty is preserved even as execution becomes automated."

Hanno Lustig describes "America's Risky Debt: What Markets See That Policymakers Don't" (Aspen Economic Strategy Group, August 2026, forthcoming in *The American Economy in a New Era*, edited by Melissa S. Kearney and Luke Pardue, https://www.economicstrategygroup.org/publication/americas-risky-debt-what-markets-see-that-policymakers-dont/). "For decades, the US has been the world's safe-asset provider, and the dollar has been the world's reserve currency. . . . Pre-2020, the safe-asset model was a good fit: Treasurys traded at a premium to close substitutes, hedged equity risk, and rallied during stress episodes. Post-2020 each of these predictions has failed. First, US Treasurys are no longer expensive compared to close substitutes like German sovereign bonds or AAA US corporate bonds, especially at longer maturities. Second, the US stock–bond correlation turned positive in 2020. US Treasury allocations no longer hedge the equity risk in an investor's portfolio. Third, there is increasing evidence that investors no longer flee to the safety of Treasurys when volatility spikes in financial markets. An example of this was the arrival of COVID-19 in March 2020, when the 10-year Treasury yield spiked by 68 basis points in just eight trading days between March 9 and March 18. . . . Bond investors increasingly question the safety of US Treasurys, and they have re-priced Treasurys as a risky claim. Central bankers and financial regulators have not adjusted: their models and financial regulation still treat government debt as unambiguously safe."

Steven Ruggles diagnoses "The pig in the python: US decennial labor flows and economic opportunity, 1910–2040" (*PNAS*, May 15, 2026, vol. 123, #20, e2601716123, https://www.pnas.org/doi/10.1073/pnas.2601716123). "We are on the verge of a fundamental transformation of the labor market. Baby boomers began retiring in large numbers during the 2010s, and immigration began declining in the same period. The number of births went down 17% between 2007 and 2024, which will

contribute to a reduction in the number of new entrants to the labor force in the 2030s. In the current moment of political turmoil, any economic predictions spanning the next 15 years are highly uncertain. We can predict with some confidence, however, the general demographic configuration of the 2040 working-age population: the births have already occurred, mortality change is ordinarily gradual, and a massive surge of immigration seems unlikely. . . . Barring revolutionary changes in the economy, the drop in labor-market competition over the coming two decades will have profound consequences. There is likely to be an unprecedented shortfall of new workers, creating strong upward pressure on wages. . . . Americans born in the 2020s might be the first cohort in a half century that earns significantly more than their parents did. Labor-force participation will probably increase as more workers are pulled into the market by higher wages, and some workers will postpone retirement."

The *Behavioral Economics Guide 2026*, edited by Alain Samson, contains two essays on frontier topics of behavioral economics (June 2026, https://www.behavioraleconomics.com/be-guide/the-behavioral-economics-guide-2026/). In "Homo Experiens: Why Behavioral Economics Needs the Life Sciences," Ulrike Malmendier writes: "To put it in stark terms, human behavior still seems rather robotic. Where previously, in the neoclassical model, it was the output of a perfectly programmed computer, in the behavioral economics model it became the output of a not-quite-so-perfect computer program, one prone to systematic bugs. And while the discovery of heuristics and biases was a genuine improvement, the resulting models remained mechanical. Once we identify the biases and model them, the program is assumed to run the same way for everyone, at every point in their lives. What you have lived through, what you have felt, what has happened to your body and brain along the way—none of that enters the model. . . . The implication is that the biological channels through which experiences affect us—stress hormones, neural rewiring, immune and endocrine responses—are real, measurable, and consequential for the economic outcomes we care about. And they are almost entirely absent from economic models, whether behavioral or not. . . . I believe that the next chapter of behavioral economics lies in a turn towards the life sciences—in taking seriously what neuroscience, psychiatry, medicine, and biology already know about human beings and how they are shaped by what they live through." In "Beyond Preferences: The Biological Foundations," Isabelle Brocas argues: "A biologically informed economics does not require economists to abandon preferences or reduce behavior to neurons and genes. Preferences remain useful reduced-form representations when behavior is stable and prediction is the objective. The point is different: when economists seek to explain heterogeneity, diagnose why interventions fail, or design policies that work across contexts and populations, it is often necessary to model the mechanisms that generate the outward pattern of choice. Biology can enter economics not only empirically, but also theoretically through models that derive observed behavior from underlying brain processes and resource constraints. This perspective clarifies what is malleable, when it is malleable, for whom, and through which policy levers. Taking biology seriously is therefore not

a detour from economic theory—it is a step toward a more explanatory and more useful science of human behavior. . . . Neuroeconomic work has been especially valuable here because it shows that value-based choice relies on a sequence of interacting computations rather than on isolated behavioral modules."

Symposia

The World Bank has published a collection of six intertwined essays on *The Global Jobs Challenge*, edited by Tommy Chrimes, M. Ayhan Kose, and Kersten Stamm (June 2026, https://www.worldbank.org/en/research/publication/global-jobs-challenge). From the "Executive summary": "Emerging market and developing economies (EMDEs) face a jobs challenge of historic proportions. Over the decade to 2035, around 1.2 billion young people in these economies are set to reach working age, the largest youth cohort the world will likely ever see. Yet global economic growth is at a multi-decade low, constraining job creation prospects. Many of these young people live in economies ill-equipped to confront the challenge. Although projections for employment over the next decade are highly sensitive to underlying assumptions, an illustrative extrapolation suggests that, of these 1.2 billion young women and men, just over 400 million would be employed in 2035, while about 300 million would not be in employment, education, or training. Jobs are at the heart of development: they provide income, purpose, hope, and dignity, while strengthening social cohesion and stability. Job creation underpins growth, poverty reduction, and shared prosperity. Creating sufficient job opportunities for this large wave of young working-age people is thus a pivotal development challenge and a potentially transformational opportunity."

The *Annual Review of Economics* (2026, https://www.annualreviews.org/content/journals/economics/browse) has published a three-paper "Symposium on the Economics of Religion." Paul Seabright contributed "The Economics of Religion: A Review." "Religion has been an integral part of economic life for centuries. Religious institutions like churches, temples, and monastic orders were economically important actors in many historical societies. Ancient Greek and Roman temples were vigorous commercial entities offering services for a price to a discriminating public. During the Middle Ages, Christian churches, Hindu and Buddhist temples, and Muslim mosques accumulated large amounts of agricultural land through bequests. By the year 750 CE, the Catholic Church owned around one-third of all the agricultural land in Western Europe as well as other assets such as buildings and slaves. The Church's wealth allowed a regular flow of resources to the poor and sick as well as the erection of churches, monasteries, cathedrals, and bishops' palaces. It also provided working capital for businesses with which the Church was more or less closely involved, from banks to pawnbroking operations, textile factories, travel agencies, breweries, and many others. . . . Some of the earliest manuals of management practice were written by and for monks and priests. . . . Given this history, it is unsurprising that Adam Smith devoted many pages in Book Five of *The Wealth of*

Nations to issues of incentives and management in religious organizations. It is more startling that for two centuries afterward, religion dropped completely off the map of the economics profession. . . . This may have been because economists followed the post-Enlightenment tendency to define religions as constituted by doctrines and beliefs. In doing so they lost sight of the fact, of which Adam Smith was well aware and which contemporary research has now placed center stage, that religious individuals engage also in important resource-consuming and resource-producing activities. . . . The modern economics of religion emphasizes that, rather than absorbing religious habits passively, religious people engage in lucid and reflective choice. To be sure, they inherit habits, norms, and values through their families, and their choices may be (sometimes severely) constrained. However, families themselves often make careful choices about how much to invest in the transmission of their cultural values, including religious values, to children. . . . Overall, religion is a major economic force in the world, both in the revenues it commands and in the human energies it mobilizes. . . . The overall message is that modern economics can help understand religion by treating religious movements as organizations that jointly supply belief, ritual, insurance, and social networking."

The *Review of Political Economy* (2026, 38:3, https://www.tandfonline.com/toc/crpe20/38/3) has published an eight-paper symposium asking: "What Have We Learned from Summers and Piketty Ten Years On?" As one example, Steven Pressman contributed "Secular Stagnation After Piketty and Keynes" (pp. 961–979). "Nonetheless, with fresh memories of the Great Depression, economists worried about secular stagnation, especially given the high levels of government debt relative to GDP. Samuelson (1948) developed a clever solution to this problem . . . With a little math, Samuelson showed that the balanced budget multiplier was equal to one . . . If an economy had an MPC of .8, raising taxes by $1 would tax away 80 cents of spending and 20 cents of savings. The former would reduce consumer spending and economic growth; the latter would not impact the economy. There would be a net spending gain of 20 cents—$1 more government spending and 80 cents less consumer spending. In Samuelson's simple model, an MPC of .8 implies a multiplier of 5. So 20 cents more net spending would lead to $1 of additional economic growth, and the balanced budget multiplier was equal to 1. . . . Samuelson's balanced budget solution works only if additional government spending goes to the middle class and the poor, who have an MPC of close to 1, while higher taxes are paid by the rich, who save a large fraction of their income and don't receive much from increased government spending. This distributional change would then change aggregate consumption. . . . More recently, Post Keynesians have taken up the mantle of changing the distribution of income in order to expand economies and have pushed for higher taxes on corporate profits."

Magnus Henrekson, Christian Sandström, and Mikael Stenkula have edited a collection of 16 essays with a European focus about *A Green Entrepreneurial State*? *Exploring the Pitfalls of Green Deals* (Springer International Studies in EntrepreneurshipVolume 49, 2026, https://link.springer.com/book/10.1007/978-3-032-15512-2). From the "Preface": "Green Deals—broad strategies that aim to

transform economies by making them climate-neutral, resource-efficient, and competitive all at once—are part of this trend. Nuclear power has been shut down prematurely in parts of Europe and replaced with intermittent energy sources such as wind and solar power. Public funds have been allocated to hydrogen initiatives, carbon capture technologies, and similar efforts to transform Western economies. By 2025, it became increasingly evident that many of these efforts were faltering. Europe's flagship battery initiative, Northvolt, ended up as the largest bankruptcy in modern Swedish history. Attempts to produce steel using hydrogen were abandoned, despite billions of euros in support from the EU and its member states. In Germany, energy prices hit record highs, and the manufacturing sector was struggling. The blackout in Spain and Portugal raised questions about grid stability in countries increasingly dependent on solar and wind power. These failures are neither small nor insignificant. They stand in stark contrast to the promises made by EU officials and renowned scholars."

Interviews

Christopher Jeffrey interviews "Mervyn King on his career and big 'mistakes.'" The subtitle reads: "The former BoE governor speaks with Christopher Jeffery about the value of transparency, lessons from the GFC [global financial crisis], the need to track broad money and tie short-term funding to haircut collateral, and danger presented by non-banks" (*Central Banking*, June 30, 2026, https://www.centralbanking.com/central-banks/monetary-policy/7976206/mervyn-king-on-his-career-and-big-mistakes, free registration required). On central bank transparency: "I remember the very first year that I joined the Bank of England full-time, in 1991, Paul Volcker came over to London . . . I said to him: 'Give me one word of advice for a new central banker.' And Volcker, who was 6 feet 7 inches tall–so he certainly looked down on me–just said: 'One word, Mervyn: "mystique".' I came to the view that we had to go in the opposite direction. And, in 1992, when Britain was forced out of the Exchange Rate Mechanism, we had to develop a new domestic framework for monetary policy. Although Britain wasn't the first to introduce inflation targeting –we followed New Zealand and Canada–we were the first major bank to push explanations of policy to the forefront." The "pawnbroker for all seasons" method of preparing for the next banking crisis: "I think the secret here is to deal with many of these problems before the crisis hits. That's why I proposed a system where banks pre-position collateral with the central bank. And the central bank says you cannot issue more short-term funding–anything that rolls over within three or four months–then the amount of money the central bank is willing to lend against the collateral that the bank has pre-positioned, allowing for the haircuts that the central bank would impose. If it's a very obscure financial instrument that we don't really understand, the central bank might impose a haircut of 90%. If it's long-term government bonds, it could be just 2% or 3%. . . . It has the great benefit that, done properly, bank runs would just disappear as a threat. I think regulation will carry on

developing in that direction. I don't expect any central bank to suddenly say they've 'seen the light', and want to adopt a 'pawnbroker for all seasons'. They'll use their own language . . . But people are making progress in that direction, and the Bank of England has gone further than anyone else."

Tyler Cowen serves as interlocutor in "Joel Mokyr on Clans, Corporations, and a Culture of Growth" (*Conversations with Tyler*, July 8, 2026, https://conversationswithtyler.com/episodes/joel-mokyr/). "[I]f you look at the world, say, around 800, at the time of Charlemagne, the difference between Europe and China isn't very large. At some point, during the Middle Ages, you can see this divergence getting started. What's happening is that, in Europe, there is more and more of a decline in the extended family or the extended kinship group, we call it clan, and instead, people get together and cooperate with other people to whom they are not related and with whom they do not share an ancestor. Whereas in China, it moves exactly in the other direction. In China, you get more and more people getting organized by their extended family. The reasons for that are fairly complex. . . . In China, there is no Catholic Church. The imperial bureaucracy is more and more in cahoots with local clans to whom they actually outsource a fair amount of the things that they were supposed to do. As you move on out of this period of the Song dynasty into later dynasties, you see this thing growing. The problem in Europe is that the nuclear family, which became the fundamental building block of society, is too small to provide local public goods. You need to cooperate with others. What emerges in Europe, and quite spontaneously, is a bunch of things that provide these local public goods that you just don't see in China. For instance, we have something called universities. We have monasteries. We have autonomous cities. All of those things are what we call corporations. What it is, is people who are not related, but what they share is not an ancestor but an objective. Guilds have one kind of objective, universities have another one, and so on and so forth. That divergence in social organization turns out, in our view, to be one of the key components of the divergence between Europe and China."

Discussion Starters

Tony Payan, Gabriela Siller Pagaza, and José Iván Rodríguez-Sánchez explore "Locked in Low Gear: Mexico's Struggling Economy" (Rice University Baker Institute for Public Policy, April 15, 2026, https://www.bakerinstitute.org/research/locked-low-gear-mexicos-struggling-economy). "Despite important advantages, such as its proximity to the United States, a still-advantageous access to U.S. markets through the U.S.-Mexico-Canada (USMCA) trade accord, and solid industrial base built over more than three decades—Mexico's economy has been consistently experiencing slow growth. . . . Between 2018 and 2024, Mexico's average growth rate was approximately 0.9% per year . . . [T]he Mexican economy seems to be locked in low gear. . . . Over half of all workers have remained in the informal economy, which is generally comprised of jobs not regulated, protected, and taxed by the

government. . . . These workers tend to be among the least productive, as they largely lack the capital, skills, and technology to integrate into the formal economy and do not have access to the formal economy's employment protections. Moreover, current informal employment levels place significant limits on Mexico's ability to raise additional revenue to expand sustainable social welfare programs and invest in infrastructure, education, science and technology, and health."

Matthew Wells tells the story of "The Bank of North America," subtitled "Before the First Bank of the United States, another, less well-known institution helped the fledgling United States find its financial footing (*Econ Focus*: Federal Reserve Bank of Richmond, Third Quarter 2026, https://www.richmondfed.org/publications/research/econ_focus/2026/q3_economic_history). "The nascent country's financial situation was a major source of uncertainty. The weak and decentralized federal government created under the 1777 Articles of Confederation, along with the states, carried a debt burden of at least $70 million, owed to domestic lenders, as well as Dutch banks and the French and Spanish governments. To put the debt in context, the colonial GDP prior to the war was about $170 million. After five years of war, that number was undoubtedly much lower, possibly making the new country's cumulative debt equal to a significant portion of its annual economic output. To make matters more complex, the U.S. currency was all but worthless thanks to the Second Continental Congress' decision to finance the war by just printing money without the necessary gold or silver specie to back it (giving rise to the expression that items of little or no value were "not worth a Continental"). . . . Hamilton is credited as the architect of the postrevolutionary U.S. financial system that allowed the country to not just survive but grow and prosper. In 1790 and 1791, he led the effort to charter the First Bank of the United States, the country's first national bank. But in the period between the end of the war and the creation of the First Bank, another smaller, far less well-known institution—the Bank of North America—served as proof of concept that addressed these initial problems and got the country pointed in the right financial direction. It was chartered by the Confederation Congress in late 1781 and opened for business in Philadelphia in January 1782."

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. Change of address notice must be received at least six weeks prior to the publication month.

Copyright © 2026 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 the 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; email: aeainfo@vanderbilt.edu.

EXECUTIVE COMMITTEE

Elected Officers and Members

President
KATHARINE G. ABRAHAM, University of Maryland

President-elect
JANICE C. EBERLY, Northwestern University

Vice Presidents
RACHEL E. KRANTON, Duke University
JAMES H. STOCK, Harvard University

Members
AMANDA E. KOWALSKI, University of Michigan
LARA D. SHORE-SHEPPARD, Williams College
ELIZABETH U. CASCIO, Dartmouth College
DAMON JONES, University of Chicago
KIMBERLY CLAUSING, University of California, Los Angeles
CHRISTOPHER TABER, University of Wisconsin–Madison

Ex Officio Members
JANET CURRIE, Princeton University
LAWRENCE F. KATZ, Harvard University

Appointed Nonvoting Members

Editor, The American Economic Review
ERZO F.P. LUTTMER, Dartmouth College

Editor, The American Economic Review: Insights
LARRY SAMUELSON, Yale University

Editor, The Journal of Economic Literature
DAVID H. ROMER, University of California, Berkeley

Editor, The Journal of Economic Perspectives
HEIDI WILLIAMS, Dartmouth College

Editor, American Economic Journal: Applied Economics
RAYMOND FISMAN, Boston University

Editor, American Economic Journal: Economic Policy
C. KIRABO JACKSON, Northwestern University

Editor, American Economic Journal: Macroeconomics
AYŞEGÜL ŞAHIN, Princeton University

Editor, American Economic Journal: Microeconomics
NAVIN KARTIK, Yale University

Secretary-Treasurer
PETER L. ROUSSEAU, Vanderbilt University

OTHER OFFICERS

Director of AEA Publication Services
ELIZABETH R. BRAUNSTEIN

Counsel
LAUREN M. GAFFNEY, Bass, Berry & Sims PLC

ADMINISTRATORS

Director of Finance
ALLISON BRIDGES

Convention Manager
REBEKAH LOFTIS

The Journal of
Economic Perspectives

Fall 2026, Volume 40, Number 4

Symposia

Economics of Higher Education

Zachary Bleemer and Jesse Rothstein,
"The Meritocratic Consensus and Stratification in Higher Education"

E. Jason Baron, Shawn Kantor, and Alexander Whalley,
"Can Government Support for Innovation Create the Next Silicon Valley?"

William G. Howell,
"Trust Violations in Research Universities"

Private Debt

Victoria Ivashina,
"The Role of Private Debt in the Financial Ecosystem"

Gregor Matvos, Tomasz Piskorski, and Amit Seru,
"Private Credit, Balance Sheets, and Financial Stability"

Articles

Susan Athey, Neil Gandal, and Tyler Moore,
"Systemic Cyber Risk: Linking Measurement, Market Incentives, and Policy"

Isaiah Andrews, Sarah Moon, and Jesse M. Shapiro,
"In Praise of the Bayesian Bootstrap"

Dean Karlan,
"The Why and How of Foreign Aid"

David J. Deming, Raffaella Sadun, and Christopher Stanton,
"The Role of Organization in Firm-Sponsored Training"

Feature

Timothy Taylor,
"Recommendations for Further Reading"

AMERICAN ECONOMIC ASSOCIATION