

*The Journal of*

*Economic Perspectives*

*A journal of the  
American Economic Association*

*Spring 2024*

# The Journal of Economic Perspectives

*A journal of the American Economic Association*

## **Editor**

Heidi Williams, Dartmouth College

## **Coeditors**

Jonathan Parker, Massachusetts Institute of Technology

Nina Pavcnik, Dartmouth College

## **Associate Editors**

Panle Jia Barwick, University of Wisconsin-Madison

Anusha Chari, University of North Carolina

Karen Clay, Carnegie Mellon University

David Deming, Harvard University

Andrea Eisfeldt, University of California at Los Angeles

Rachel Glennerster, University of Chicago

Louis Kaplow, Harvard University

Imran Rasul, University College London

Juan Carlos Serrato, Duke University

Lise Vesterlund, University of Pittsburgh

Jeffrey Wooldridge, Michigan State University

## **Data Editor**

Lars Vilhuber

## **Managing Editor**

Timothy Taylor

## **Assistant Managing Editor**

Bradley Waldruff

---

### *Editorial offices:*

Journal of Economic Perspectives

*American Economic Association Publications*

2403 Sidney St., #260

Pittsburgh, PA 15203

email: [jep@aeapubs.org](mailto:jep@aeapubs.org)

---

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

Copyright © 2024 by the American Economic Association; All Rights Reserved.

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

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

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

*THE JOURNAL OF ECONOMIC PERSPECTIVES* (ISSN 0895-3309), Spring 2024, Vol. 38, No. 2. 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 38 • Number 2 • Spring 2024*

---

**Symposia**

***How Research Informs Policy Analysis***

- Staff of the Congressional Budget Office, “How Economists Could Help Inform Economic and Budget Analysis Used by the US Congress” . . . . 3
- Cass R. Sunstein, “The Economic Constitution of the United States” . . . . . 25
- Wendy Edelberg and Greg Feldberg, “The Financial Crisis Inquiry Commission and Economic Research” . . . . . 43
- Emily Oehlsen, “Philanthropic Cause Prioritization” . . . . . 63

***Labor Market and Macroeconomics***

- Gadi Barlevy, R. Jason Faberman, Bart Hobijn, and Aşegül Şahin, “The Shifting Reasons for Beveridge Curve Shifts” . . . . . 83
- Loukas Karabarbounis, “Perspectives on the Labor Share” . . . . . 107
- Richard Rogerson, “Why Labor Supply Matters for Macroeconomics” . . . . . 137
- Marianna Kudlyak, “How Cyclical Is the User Cost of Labor?” . . . . . 159

***Privacy Protection and Government Data***

- Claire McKay Bowen, “Government Data of the People, by the People, for the People: Navigating Citizen Privacy Concerns” . . . . . 181
- Steven Ruggles, “When Privacy Protection Goes Wrong: How and Why the 2020 Census Confidentiality Program Failed” . . . . . 201

**Articles**

- Emmanuel Saez, “Gabriel Zucman: Winner of the 2023 Clark Medal” . . . . . 227

**Features**

- Timothy Taylor, “Recommendations for Further Reading” . . . . . 245

## **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**

Kerwin Charles, Yale University

Cecilia Conrad, Lever for Change

Gary Hoover, Tulane University

Jeffrey Kling, Congressional Budget Office

Dylan Matthews, Vox.com

Emily Oster, Brown University

Catherine Rampell, Washington Post

Otis Reid, Open Philanthropy

Ted Rosenbaum, Federal Trade Commission

Lucie Schmidt, Smith College

David Wessel, Brookings Institution

# How Economists Could Help Inform Economic and Budget Analysis Used by the US Congress

Staff of the Congressional Budget Office

**T**he US Congress uses economic and budgetary projections, cost estimates for proposed legislation, and other analyses provided by the Congressional Budget Office (CBO) as part of its legislative process. CBO makes assessments based on an understanding of federal programs and revenue sources, reading the relevant research literature, analysis of data, and consultation with outside experts—and often relies on economic research.

This article begins with a discussion of the role of the Congressional Budget Office and then discusses how economists could conduct research that would help inform the Congress by improving the quality of the analysis and parameter

■ *The author of this article for citation purposes should be “Staff of the Congressional Budget Office.” The article includes contributions from the following CBO staff, listed here in alphabetical order: Adebayo Adedeji (formerly of CBO), Aaron Betz, Dorian Carloni, Nicholas Chase, Carrie H. Colla (formerly of CBO), Daniel Crown, Molly Dahl, Richard DeKaser, Elizabeth Cove Delisle, Devrim Demirel, Noelia Duchovny, Justin Falk, Michael Falkenheim, Ann E. Futrell, Sebastien Gay, Ron Gecan, Heidi Golding (formerly of CBO), Bilal Habib, Rebecca Heller, Evan Herrnstadt, Justin Humphrey, Nadia Karamcheva, Edward G. Keating, Joseph Kile, Wendy Kiska, Jeffrey R. Kling, Leah Koestner, Eric J. Labs, Mark Lasky, Junghoon Lee, Chandler Lester, Sarah Masi, John McClelland, Noah Meyerson, David Mosher, Jaeger Nelson, Xiaotong Niu, Daria Pelech, Jeffrey Perry (formerly of CBO), Joseph Rosenberg, Molly Saunders-Scott, Jeffrey Schafer, Chad Shirley, Emily Stern, William Swanson, Julie Topoleski, and Chapin White. Kling is the corresponding author at [Jeffrey.Kling@cbo.gov](mailto:Jeffrey.Kling@cbo.gov).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.2.3>.

estimates that CBO uses. It gives overall context and specific examples in seven areas: credit and insurance, energy and the environment, health, labor, macroeconomics, national security, and taxes and transfers. The examples are intended to be illustrative of some current priorities, not a comprehensive list.

## **The Work of the Congressional Budget Office**

Since 1975, the Congressional Budget Office has produced nonpartisan and independent analyses of budgetary and economic issues to support the Congressional budget process. Each year, the agency produces economic forecasts and baseline projections of revenues and spending that generally follow current law as well as hundreds of cost estimates for many types of legislation (CBO 2023b). The staff preparing budgetary analyses draw on contributions by other economists and researchers who also produce dozens of related analyses for the Congress to help make CBO's work transparent and to respond to requests on topics ranging from income distribution to weapon systems.

Based on the laws that established the Congressional Budget Office, its analysis of proposed legislation ultimately focuses on federal budgetary implications. CBO primarily focuses on estimating effects on outlays. The staff of the Joint Committee on Taxation (JCT) in the Congress produces cost estimates of legislative changes to income, estate and gift, excise, and payroll taxes. For analyses of policies affecting both outlays and revenues, CBO collaborates with JCT.

When tasked with analyzing major legislative proposals that would affect health insurance choices for people under age 65, for example, the Congressional Budget Office focuses on estimating the effects on coverage, premiums, and federal spending, and the staff of the Joint Committee on Taxation estimates the tax-related budgetary effects. CBO and JCT embark on the following four-step process—using the same steps that are employed for many other types of proposals—to develop a cost estimate that reflects the middle of the distribution of potential outcomes (CBO 2020a)

(1) *Review the proposal.* The agencies read the draft proposals, often beginning with informal ones as ideas take shape and ending with a final version of the legislative language. They analyze the policy specifications in the drafts, clarify any ambiguities, and identify how the proposal would change federal laws and interact with state laws.

(2) *Develop a modeling strategy.* The agencies identify the proposal's potential key effects on individuals' and employers' health insurance decisions and the relevant effects on insurers' and states' decisions about coverage and benefits—along with effects on health care providers. They determine the timing of the effects, such as altering or starting programs or changing state laws or regulations, and review existing research, particularly on similar policies or programs. They consult with outside experts, including insurance commissioners, actuaries, benefit consultants, and researchers. They decide which models to use—ranging in complexity from

spreadsheets to simulations using thousands of lines of computer code—whether their capabilities need to be extended, and how to translate the information gathered into inputs to the models.

(3) *Model the effects of the proposal.* Having included inputs for insurers' and state governments' decisions about coverage and benefits, the Congressional Budget Office models individuals' and employers' health insurance decisions over the projection period using a microsimulation model (CBO 2021g). CBO estimates spending for Medicaid, the Children's Health Insurance Program, and other programs using program-specific models. The staff of the Joint Committee on Taxation estimates federal revenues using its tax models.

(4) *Review and write about the estimate.* The agencies review the models' output for analytical soundness and objectivity and assess the main sources of uncertainty and possible alternative outcomes. They write the cost estimate, review it for clarity, and then publish it.

The Congressional Budget Office aims to make the basis of its estimates transparent. In that effort, the agency often reports estimates of effects that are building blocks of budgetary analysis and of independent interest to policymakers. For example, CBO often reports estimated changes in health insurance or premiums for proposals that affect health insurance coverage or prices because an analysis of those outcomes is necessary to complete the estimate (for example, CBO 2022b). Lawmakers can put their own weights on different elements of such information when making their decisions about whether or not to support a particular proposal. CBO neither undertakes cost-benefit analyses—which would involve using particular weights to aggregate information—nor makes policy recommendations.

The analysis of the Congressional Budget Office generally focuses on projections over the next ten years, because the Congress's budget process tracks goals for spending and revenues over that period of time as specified in law (CBO 2023c). When feasible, CBO provides supplemental information about longer-term budgetary effects of various policies, such as for spending on physical infrastructure (CBO 2021f) and Social Security (CBO 2022e). The budget process does use present values of future cash flows beyond ten years for a few programs, like credit programs that have loans outstanding beyond that horizon. CBO estimates the net lifetime costs of new loans and loan guarantees in the year of issuance (CBO 2023g). Evaluations of the extent to which laws have achieved their goals in the past is typically undertaken by outside researchers, or within the government by agencies like the Council of Economic Advisers, the Government Accountability Office, or more focused agencies like the Environmental Protection Agency.

The Congressional Budget Office's cost estimates for legislation account for changes in the total output of the economy—through work known as dynamic analysis—in limited circumstances. According to long-standing practice, CBO's conventional cost estimates reflect the expectation that total economic output measured in current-year dollars would not change. In its estimates of the effects of increasing the minimum wage, for instance, CBO (2023k) reported results using both conventional and dynamic approaches. Although some major legislation

can have significant macroeconomic consequences—say, because it would affect the labor supply or private investment—most does not. The Congress has directed CBO to include dynamic analysis in its cost estimates, if practicable, when the gross budgetary effects of a bill would equal or exceed 0.25 percent of the economy’s output in any year. Such estimates are complicated and often time-consuming, so they are difficult to prepare if legislation is moving quickly. Even though most cost estimates do not reflect the macroeconomic effects of a particular bill, they do reflect (when relevant) the effects that changes in policy might have on people’s behavior that would, in turn, produce budgetary effects. For example, an estimate might account for the likelihood that people would take up a particular government benefit under a new law, the possibility that farmers would change what and how much they grow in response to a change in agriculture programs, or the ways that businesses might adjust their operations in response to a particular subsidy.

## **Credit and Insurance**

The Congressional Budget Office regularly provides information to the Congress about the effects of proposed policies that would modify federal credit and insurance programs including student loans and pension insurance.

### **How Would Borrowers Respond to Major Changes in Repayment Plans for Student Loans?**

The cost of the federal student loan program depends in part on the details of repayment plans. Starting in 2009, income-driven repayment plans have expanded: in these plans, borrowers pay a fixed percentage of their income to the student loan program, either until they have paid back the loan or until they have made a certain number of payments. However, the estimates from the Congressional Budget Office of income-driven repayment initially tended to underestimate the number of borrowers that would use those plans and the degree to which the plans would be adversely selected by borrowers with high loan balances and low income. Estimates of repayment plan choice would have benefited from empirical studies examining specific segments of the heterogeneous population of borrowers that would be affected by such policies (such as a shorter repayment period or a lower monthly payment amount) or from surveys on borrowers’ preferences. For a recent survey of the literature on student lending, see Yannelis and Tracey (2022).

A new income-driven repayment plan was finalized in July 2023, which the Congressional Budget Office has estimated will increase the cost of the federal student loan program by about \$230 billion, on a net-present-value basis, over the 2023–2033 period (CBO 2023e). Because the new plan reduced borrowers’ costs, CBO expected it to be the most popular repayment plan option and that amounts borrowed would increase under the plan. But when CBO was preparing its micro-simulation model of student borrowers (Karamcheva, Perry, and Yannelis 2020) to estimate the additional enrollment in income-dependent repayment plans, the



agency did not find research that was directly relevant to projecting a borrower's repayment plan choice or borrowing amounts. Although some studies had examined the effects of expanding the availability of student loan credit (Black, Turner, and Denning 2023; Kelchen 2019) and increasing borrowing limits (Kargar and Mann 2023; Lucca, Nadauld, and Shen 2019), there was no relevant research examining how changes in loan parameters—and in parameters of income-driven repayment plans in particular—would affect students' propensity to borrow, the amount students borrow, or how institutions might respond. Research in those areas could help inform CBO's baseline projections of student borrowing over the next ten years and, in turn, its projections of the cost of the student loan program.

### **How Would Sponsors of Pension Plans Respond to Changes in Government Pension Insurance?**

The government provides pension insurance for the “defined benefit” pension plans of private-sector companies through the Pension Benefit Guaranty Corporation (PBGC), which has experienced funding difficulties in the past two decades. In 2005, PBGC estimated that the single-employer program faced a net shortfall of \$23 billion. The Pension Protection Act of 2006 established more rigorous funding requirements for covered plans, and PBGC now projects that the single-employer program has a positive net position of \$45 billion and will remain solvent indefinitely. The multiemployer program did not face significant exposure at the time, but developed a shortfall of \$65 billion by 2019 and faced insolvency within a few years. The American Rescue Plan of 2021 provided about \$80 billion of special financial assistance to the most financially troubled plans, but left the structure of the program largely unchanged.

Cash flows for pension insurance are tracked in the federal budget: insurance premiums from companies with such pension funds are recorded when received; claims from bankrupt funds are recorded as outlays when paid. The Congressional Budget Office and the staff of the Joint Committee on Taxation are tasked with analyzing the budgetary implications of federal policies that affect contribution requirements, premium payments, investment restrictions, and other funding rules related to pensions.

Over the past 25 years, large employers have reduced their use of defined-benefit pension plans that are covered by pension insurance and shifted to individual-level “defined contribution” plans that are not covered. This change has often been carried out by freezing the defined benefit accruals for workers and/or closing the plans to new hires. One study found that freezing pension plans saves 13.5 percent of the present value of payroll in the long run (Rauh, Stefanescu, and Zeldes 2020). Additional research that identifies the factors that are likely to lead to plan freezes, as well as the conditions that encourage plans' sponsors to retain defined-benefit pension plans, would be useful; for example, legislative proposals that would increase the number of plan freezes could reduce premium receipts for pension insurance provided by the federal government.

Similarly, little research is available on the incentives for participating employers to withdraw from multiemployer pension plans; that is, plans in which the defined benefit is run in common across employers in the same industry—thus allowing workers who switch employers within the industry to continue accruing pension benefits. At the Congressional Budget Office, modeling has focused on mass withdrawals from multiemployer plans, because they have been most closely related to insurance claims (Kiska, Levine, and Moore 2017). However, partial withdrawals where only a few firms leave a multiemployer plan have significantly affected the level of plan funding and often impose financial strain on the employers remaining in the plans. Additional research on the factors that encourage employers to remain in or withdraw from pension plans would help CBO to better estimate participation (which affects premium receipts) and plans' future financial outcomes (which affect future federal outlays).

Funding for both single-employer and multiemployer pension plans is sensitive to price volatility in financial markets in which pension assets are invested in risky securities. Moreover, federal insurance for private pension plans may encourage risk-taking. For example, one study has found that underpricing pension insurance encourages plans' sponsors to invest in risky assets (Love, Smith, and Wilcox 2011). Additional research on the factors that influence how plan administrators manage risk in response to both the pricing and the level of federal insurance would help the Congressional Budget Office estimate the government's future outlays. When lawmakers consider pension reforms, they may also be interested in the optimal pricing and structure of pension insurance, as well as its relationship with investment policy.

## **Energy and the Environment**

Federal regulations and permitting requirements can affect the amount and composition of energy produced and used, prices of different sources of energy, and carbon dioxide (CO<sub>2</sub>) emissions from using energy. The federal government also provides funds to prepare for disasters and reduce the resulting damage they cause. The Congressional Budget Office is particularly on the lookout for research in these two areas.

### **How Would Changes to the Federal Permitting Process Affect Energy Markets, CO<sub>2</sub> Emissions, and the Macroeconomy?**

Congress often considers legislation intended to shorten the time it takes to obtain the variety of federal, state, local, or, in some cases, tribal permits or approvals required to develop infrastructure to produce and deliver supplies of energy—wind and solar generating facilities or oil and gas pipelines, among others. The National Environmental Policy Act additionally requires federal agencies to consider the potential environmental effects of their decisions when granting project approvals (Congressional Research Service 2022; Council on Environmental Quality 2021).

Although some projects qualify for abbreviated review under this legislation, others are subject to extensive environmental study—particularly infrastructure that covers long distances and crosses multiple jurisdictions, such as natural gas pipelines and electric transmission lines. Some industry analysts and lawmakers have raised concerns that the time required for those environmental reviews has grown lengthy (Dourado and Smith 2020).

Again, the Congressional Budget Office is responsible for estimating the effects of proposed legislation on the federal budget (for example, CBO 2023f, 2019b, 2018b) and for incorporating those effects in its baseline projections if the legislation is enacted. For example, shortening the period for preparing and finalizing an environmental impact study under the National Environmental Policy Act could affect the budget through at least two channels. First, reducing permitting timelines would probably accelerate project development and boost royalty payments on oil and gas produced on federal lands, for example. The increase in production would come from reducing the amount of time projects await federal approvals and from a greater propensity for developers to invest in new projects. Second, increases in aggregate productivity from greater capital investment and from lower costs of energy would generate broader macroeconomic effects that, in turn, would increase tax revenues.

Analyzing the effects of regulations on businesses and the economy has been an active line of research (Coffey, McLaughlin, and Peretto 2020; Dawson and Seater 2013; Djankov, McLiesh, and Ramalho 2006). However, little research has considered the expected economic and environmental effects in energy-producing sectors from changes to permitting or to environmental reviews under the National Environmental Policy Act or other laws. A recent report on the nonbudgetary effects of charging the oil and gas industry for methane emissions (CBO 2022d) made use of the available, albeit limited, research in that area.

Some research has considered federal environmental protections as a whole but has not identified the relative importance of particular requirements for environmental protection (Lewis 2019). Other studies have focused on specific environmental protections but not the role of environmental review under the National Environmental Policy Act (Ryan 2012; Greenstone 2002). Research that evaluates how changes to the federal permitting process and related environmental reviews would affect domestic energy sectors, CO<sub>2</sub> emissions, and the macroeconomy could enhance the Congressional Budget Office's analysis.

### **How Would Changes in Federal Spending on Climate Change Adaptation Affect Damage?**

When hurricanes and other natural disasters occur, lawmakers often fund construction projects to better adapt to climate change and provide financial assistance to people affected. Over the 2005–2021 period, appropriations for the Army Corps of Engineers averaged about \$9 billion a year (CBO 2022a). Over the 1992–2021 period, appropriations for the federal Disaster Relief Fund averaged about \$13 billion (CBO 2022c). The Congressional Budget Office is also often asked to

provide the Congress with information about how increased federal spending on measures to adapt to climate change, such as improving levees or elevating structures, might reduce the amount of damage from flooding or from other effects of the changing climate (CBO 2023h, 2019c, 2016).

In the past, the Congressional Budget Office has drawn upon an analysis of a small sample of adaptation projects (Multihazard Mitigation Council 2005) to estimate the potential cost savings attributable to spending on the Federal Emergency Management Agency's Pre-Disaster Mitigation Program (CBO 2007). A recent update of the underlying analysis did not appreciably increase the size or representativeness of the sample (National Institute of Building Sciences 2019).

The research on adaptation to climate change generally falls into three categories. First, some studies directly estimate the relationship between historical spending and subsequent disaster damage (for example, Davlasheridze and Miao 2021). One limitation of those studies is that the estimate might not reflect future savings if climate change and economic development substantially altered the earlier risks of damage. Second, other studies estimate how the presence of infrastructure intended to mitigate damage from disasters affects property values (Kelly and Molina 2023; Bradt and Aldy 2023). Those studies can reflect expectations about amounts of damage in the future, but the results will be biased to the extent that people have imperfect information or misperceive risk. Third, studies can rely on engineering-based models of disaster damage rather than econometric estimates (Neumann et al. 2021; Wobus et al. 2021; Johnson et al. 2019; Aerts et al. 2014). For example, those studies can simulate many years of hypothetical flooding rather than relying on a small sample of historical disasters. However, such models are subject to considerable uncertainty. Also, they do not typically account for behavioral responses by households, businesses, and state and local governments, which could increase risk or reduce complementary spending and affect the net reduction in damage.

The Congressional Budget Office is conducting research on the effects of federal spending on climate change adaptation, and that work would be enhanced by additional research that extends, compares, and combines different approaches to help fill gaps in the literature. Studies that cover more federal disaster mitigation programs, additional kinds of disasters, and federal programs that promote adapting to climate impacts beyond property damage (for instance, funding research on developing climate-resilient crops) would be particularly useful.

## **Health Care**

When the Congressional Budget Office projects federal subsidies for health care, it begins by estimating enrollment in various forms of health insurance coverage, premiums for that coverage, and prices and the use of medical items and services. Those projections, in turn, require estimates of certain parameters to construct equilibrium prices, premiums, and use of services. CBO relies on its

own analyses and on estimates from the relevant research literature to inform its estimates of those parameters. For instance, after reviewing evidence about how changes in health care providers' payment rates affect the supply of their services, CBO updated how it estimates supply responses. At present, two subjects of particular interest are how health care providers would respond to changes in federal policies and the market for long-term services and supports.

### **How Would Health Care Providers Respond to Shocks to Revenues or Costs?**

The Congressional Budget Office frequently analyzes policies affecting health care providers' payment rates or their costs of providing care (CBO 2022f). For example, policies can directly affect providers' revenues by changing the administered prices for services covered by Medicare. Alternately, they can affect the amounts paid to providers by commercial insurers by changing the bargaining leverage between them. The prices paid to providers by commercial insurers have a large effect on the federal budget: if prices paid for hospitals' and physicians' services were reduced by 1 percent, for instance, federal subsidies for health insurance premiums would be reduced by \$4.4 billion for employment-based insurance in 2032, CBO estimates.

There is a long-standing concern that payment changes—particularly cuts—could impact the availability of care. However, providers have many ways to adjust to payment changes. For example, reducing the volume of care they provide, particularly for elective procedures (Clemens and Gottlieb 2014); reducing staffing (Wu and Shen 2014); or changing how they code diagnoses (Dafny 2005). A better understanding of how adaptable providers' cost structures are, the major components of their fixed and variable costs, and how they respond to changes in revenues would enhance analysis by the Congressional Budget Office of current policies and legislative proposals.

The Congressional Budget Office also evaluates policies that affect providers' administrative burdens, such as streamlining requirements to obtain prior authorization for medical services. Based on evidence from Curto et al. (2019) and Dunn et al. (2021), the agency estimates that lessening a provider's administrative burden would tend to increase the amount of care the provider delivers, thereby increasing federal spending. Additional evidence would be helpful in refining those and other estimates.

### **How Would Changes in Medicaid's Benefit for Long-Term Services and Supports Affect the Federal Budget?**

Medicaid is the predominant payer for long-term services and supports, which consist of health care and related services to help people who have functional or cognitive limitations in performing routine daily activities over an extended period. Roughly 65 percent of total national spending for these services are paid by the federal government. In CBO's projections, by 2033, federal Medicaid spending on home and community-based services reaches \$116 billion, and such spending on institutional care is an additional \$53 billion (CBO 2023j). Those projections reflect

estimates of enrollment in Medicaid, growth in payment rates for providers of long-term services and supports, and an expectation of a continued shift in the delivery of care from institutions to community settings.

Policy changes, such as expanded Medicaid eligibility, can affect the demand for care in nursing homes (Van Houtven et al. 2020; Grabowski and Gruber 2007). Analyses by the Congressional Budget Office could benefit from additional research on the population with needs for long-term services and supports and, specifically, how changes in federal policy would affect the share of people using institutional care or home and community-based services.

The Congressional Budget Office uses several models to estimate how changes in Medicaid eligibility, benefits, or payment rates would affect the federal budget. In addition, the agency is developing a microsimulation model to better represent the distribution of individual responses to policy changes—instead of an approach focused on average responses—and important relationships between key variables, such as income and the use of long-term services and supports (Goda, Golberstein, and Grabowski 2011). Depending on the details of the policy being analyzed, CBO’s models may account for the substitutability of paid and unpaid care, changes from private or state funding to federal funding, the supply of workers, and whether changes in the use of long-term services and supports affect the use of medical services, among other factors. Additional research on those topics would improve CBO’s estimates of the effects of policies in this area.

## **Labor Markets**

The Congressional Budget Office provides the Congress with information about the effects of proposed income support policies on outcomes in the labor market (for example, CBO 2021c, 2021e). The agency also estimates the budgetary and (when relevant) macroeconomic effects of legislative proposals in these areas (for example, CBO 2021h) and incorporates such effects into its baseline projections for legislation that is enacted. To help inform such work, CBO is on the lookout for new research on various topics in the area of labor, including immigration and child care.

### **How Does Immigration Affect Productivity?**

In 2013, the Congressional Budget Office analyzed a large immigration reform bill (CBO 2013). For that analysis, the agency projected the legislation’s direct effects on the size of the US population, employment, and taxable compensation—and then incorporated those projections into its cost estimate. Since then, CBO has continued developing additional capacity to analyze a wider range of effects of changes in immigration policy. For example, the agency uses its macroeconomic models (discussed in the next section) to estimate changes in the income earned by capital, the rate of return on capital (and therefore the interest rates on government debt), and the differences in wages for workers with different skills. Recently,

leveraging those modeling tools, CBO has incorporated an increase in population, mostly because of higher net immigration, into its macroeconomic forecast (CBO 2024).

Of particular interest is how immigration policy can affect the productivity of labor and capital; that is, how immigration policy can affect the education, work experience, and other skills of immigrants, along with on how those immigrants affect other workers, the allocation of capital, and technological progress. Relatively few studies have estimated the effect of immigration on total factor productivity (Aleksynska and Tritah 2015; Ortega and Peri 2014a, 2014b, 2009), and only two have focused specifically on the United States (Peri 2012; Prato 2022).

Further research on two aspects of immigration's effect on productivity could enhance the Congressional Budget Office's analysis: more information about how the effect of immigration on productivity varies depending on the skill composition of immigrants; and evidence about the timing of how any effects of immigration on productivity will occur. Of course, how immigration affects productivity would also inform CBO's modeling of immigration's broader economic effects.

### **How Would Changes in Federal Funding for Child Care Affect Families with Children and the Child Care Industry?**

For analysis of legislative proposals related to child care, the Congressional Budget Office relies on published empirical evidence and experts' opinions, including academic researchers, state administrators of child care subsidies, and child care providers (CBO 2021d). CBO's analysts seek to assess how changes in federal funding for child care, including early childhood education, would affect families' choices in the labor market as well as the demand for and supply of child care. The effects can be economic (labor force participation of parents), distributional, and budgetary.

On the demand side, researchers have studied how changes in the availability of child care affect parents' involvement in the labor force (Li 2020; Cascio and Schanzenbach 2013; Fitzpatrick 2010). Studies of subsidized child care have focused on low-income families (Michalopoulos 2010; Berger and Black 1992). The Congressional Budget Office could benefit from research that demonstrates how those subsidies affect the labor supply of middle- or higher-income families. Also, CBO welcomes input on the take-up rates of child care subsidies among eligible families, on the extent to which families shift between unpaid care and paid care, and on the ways that large shifts in demand would affect the cost of care for families (Borowsky et al. 2022).

On the supply side, studies of the wage elasticity of child care workers and the speed at which the supply of child care (both the physical infrastructure and labor supply) would increase in response to a shift in demand could be useful in estimating the budgetary and economic effects of various policy proposals. Research on states' involvement in child care (GAO 2023) also could enhance the Congressional Budget Office's analyses. For instance, such research could shed light on how states

used the additional federal child care funding they received during the coronavirus pandemic, as well as their responses when that funding expired.

## **Macroeconomics**

A team of analysts at the Congressional Budget Office regularly prepares forecasts of key economic variables, including output, income, employment, inflation, and interest rates. The agency also evaluates the macroeconomic effects of proposed or enacted legislation. The team closely follows economic developments and data, consults with experts within and outside the agency, and uses several models, including a large-scale macroeconometric model (Arnold 2018). CBO uses a suite of models to analyze the short- and longer-term economic effects of changes in fiscal policy. In its view, fiscal policy affects the economy in the short term mainly by altering the aggregate demand for goods and services. To analyze those effects, CBO considers empirical evidence about how households, businesses, and federal, state, and local governments would respond to changes in certain policies. CBO also uses structural models that describe how policy changes would affect economic output, employment, interest rates, inflation, and other macroeconomic variables (Lasky 2022; CBO 2020b, 2014). To analyze the longer-term economic effects of fiscal policy stemming from changes in national saving, people's incentives to work and save, and businesses' incentives to invest, the agency generally uses a Solow-type growth model (CBO 2021b) and a life-cycle growth model (CBO 2019a; Reichling and Nishiyama 2015). CBO also uses dynamic general-equilibrium models and vector autoregression models to assess specific aspects of policies and the uncertainty of the economic effects of policy changes. At present, CBO is particularly on the lookout for new research on trends in productivity growth and on interest rates on Treasury securities.

### **How Will Future Rates of Productivity Growth Differ from Those in the Past?**

In the long-term economic projections from the Congressional Budget Office, growth of total factor productivity—referred to here as just “productivity”—accounts for more than half of the growth of real gross domestic product. The agency draws on academic research to assess trends in productivity growth and to estimate the effects of policy changes on productivity. That research has developed methods to assess how the trend growth rate of productivity varies across business cycles and changed during the coronavirus pandemic; whether productivity growth is additive or geometric; and the contributions to growth from factors such as workers' average educational attainment, federal investment, and climate change (Fernald and Li 2022; Philippon 2022; CBO 2021e; Bom and Ligthart 2014). Further research that identifies how those lessons from the past can be applied to projections of future productivity growth could enhance CBO's projections of economic growth, tax revenues, and government spending.



### **How Will Future Interest Rates Differ from Those in the Past?**

The real rate of return on safe assets—like Treasury securities—has declined considerably in recent decades (Rachel and Smith 2017). Researchers have examined the role of several factors in the dynamics of the returns on safe and risky assets: labor force growth, private domestic and foreign saving rates, total factor productivity growth, the debt-to-GDP ratio, risk premiums, and capital's share of income (Gamber 2020). Better understanding the magnitude and persistence of each factor and also of the linkages between the rates of return on investments in risky assets (such as physical capital) and Treasury securities would enhance the framework that the Congressional Budget Office uses to project interest rates—especially rates 5 to 30 years in the future—and net interest payments on federal debt as debt is projected to rise substantially in the future.

### **National Security**

The US Department of Defense received about \$850 billion in funding in 2023. With those funds, it hires personnel (members of the military and civilian employees) and purchases a variety of goods and services from private-sector companies—ranging from ordinary office supplies to highly complex weapon systems. Some current subjects of particular interest to the Congressional Budget Office include the consequences of the military's extensive use of in-kind compensation and the causes of, and future trends in, sector-specific inflation.

### **What Are Some Implications of the Military's Compensation System?**

The military compensation system is very complex. The Department of Defense uses a combination of methods to attract and retain members of the military and to boost productivity that is unique in the labor market: binding contracts for service commitments, bonuses, annuity payments, compensation on the basis of family status, and in-kind compensation. Members in some high-demand occupations may be eligible for reenlistment bonuses. Service members receive a larger housing allowance when they have dependents, and they receive extensive medical care and access to subsidized food stores and recreational facilities. Upon reaching 20 years of service, a member of the military becomes eligible to receive an annuity (after leaving the service) that generally amounts to between 40 percent and 50 percent of their annual basic pay, indexed for inflation.

Analysis of the military's compensation system by the Congressional Budget Office would be enhanced by more information about the consequences of in-kind compensation, which might build on existing research (Patterson, Petkun, and Skimmyhorn 2019). For example, how much do military-aged civilians value direct cash compensation compared with in-kind compensation, such as health care benefits? What is known about the selection effects of different types of compensation—for example, how do enhanced health insurance coverage, generous retirement

benefits, or subsidized childcare, gym memberships, and wellness programs affect employee attraction and retention?

### **By How Much and for How Long Can Inflation in Some Sectors Exceed Overall Inflation?**

Government acquisition of complex weapon systems differs considerably from the way most goods in the economy are purchased. A small number of companies produce such weapons in a capital-intensive industry with high barriers to entry. There is no private-sector demand for such weapons in the United States, sales to foreign governments are prohibited without government approval, and in many cases the Department of Defense cannot credibly enforce fixed-price contracts.

Naval vessels are a canonical example of a product in a monopsony—a market with the Department of Defense as the single buyer. Moreover, the Navy’s ships are manufactured by a small number of private-sector companies. Each year, the Navy submits a report to the Congress about its planned procurement of ships over the next 30 years. The Congressional Budget Office has a legislative mandate to analyze the costs of those plans (for example, CBO 2023a).

Estimates of future shipbuilding costs by the Congressional Budget Office have been consistently higher than the Navy’s. That is partly because, unlike the Navy, CBO projects that the costs of labor and materials in the shipbuilding industry will continue to grow at a faster rate than prices in the economy as a whole—as they have over the past several decades. Projections of unending supernormal cost growth in a specific sector of the economy would seem to be out of equilibrium; but such pessimistic medium-term projections of shipbuilding costs have been borne out heretofore. Other parts of the broader economy, such as hospital services and college tuition, have also experienced price increases far above the economy-wide average for many years (Perry 2022). It is unclear how long such supernormal price increases should be expected to continue.

Analysis of the likely costs of the Navy’s shipbuilding plans could be enhanced by research on fundamental questions about sector-specific inflation in consumer and producer prices. In other economic sectors that have experienced supernormal price increases for long periods, what caused the increases? Are sectors with limited competitive pressures especially prone to supernormal inflation? In cases in which supernormal, sector-specific inflation ended, what caused it to end?

## **Taxes and Transfers**

The Congressional Budget Office regularly provides the Congress with information about the ways that the government’s tax and transfer system affects the distribution of household income (for example, CBO 2023i). That analysis is built on the models and data underlying the agency’s baseline projections of revenues and spending (CBO 2023d). The method CBO uses for projecting each revenue source in its baseline varies and depends on available information. For example,

individual income taxes are projected using a microsimulation model to project the effects of tax rules using a detailed sample of tax filers (CBO 2018a), whereas revenues from the corporate income tax are projected using more aggregate methods (CBO 2023i). In this area, CBO is on the lookout for new research that would enhance its analysis of taxes and transfers, particularly including research related to distributional analysis across households and the effects of the tax system on the legal structure of businesses.

### **How Do Changes in Federal Policy Affect Different Households?**

To estimate the effects of policy changes on households along the income distribution, the Congressional Budget Office allocates to households the net dollar value of the resource costs incurred by the federal government (Habib and Heller 2022). For cash transfers and most taxes, CBO allocates the taxes and transfers according to the dollar value that households receive or pay. In other cases, including when a firm interacts with the government, CBO relies on evidence in the economic literature (Gravelle 2010, 2011) to determine how to allocate the resource costs of the government's actions. For example, the agency allocates corporate income taxes to households in proportion to their capital income (75 percent) and labor income (25 percent). For in-kind transfer programs, such as Medicaid and Medicare, CBO currently allocates 100 percent of the resource cost to people receiving coverage. For public goods and other broad-based government spending, CBO has previously allocated the resources to households by using a combination of two methods: allocating in proportion to each household's share of the population, and in proportion to each household's share of total income.

This distributional analysis differs from an examination of economic incidence, which focuses on how prices and quantities of goods and services change when government revenues or spending change and assesses who is affected by those changes and their valuation of the change. Unlike that kind of analysis, the agency's distributional allocation of resource costs is an extension of budgetary accounting that focuses on the government's revenue collections, costs of providing a good or service, and who is affected.

Research that provides more information about the impact of federal policies on households' net resources would be useful. One example is research on how to allocate resource costs for Medicaid across households. Based on a randomized experiment from expanding Medicaid coverage in the state of Oregon, Finkelstein, Hendren, and Luttmer (2019) found that only about 40 percent of the resource cost of the program accrued to people who received coverage, with the remainder accruing to providers and other parties. It is unclear whether the results from that study apply broadly to the entire Medicaid program; but, given the program's size, even modest differences in the allocation could have significant effects on the pattern and trends of household income, especially among lower-income groups. Additional research that provides estimates of the distributional effects of Medicaid and also of other health-related spending, especially in different institutional settings, would be useful.

The distributional effects of policies may differ depending on the time horizon considered or the overall state of the economy (Saez and Zucman 2023). Allocations by the Congressional Budget Office generally do not reflect those differences. For the employer share of the Social Security and Medicare payroll tax, for example, the agency has estimated that employers make adjustments to how much is passed along to employees over time, but the average amount of time they take to make that adjustment is unclear (Carloni 2021). CBO generally focuses on the long-term effects of policy changes when assessing how to make allocations—not only for the sake of simplicity and tractability, but also because distributional effects can be compared consistently with the baseline distribution of household income and other policy changes in the long term. However, Congress may be particularly interested in differences in the short- and long-term effects of policies or in the effects of policies that differ depending on the overall state of the economy. Additional research along those dimensions would allow CBO to better estimate such effects.

### **How Do Federal Taxes Affect the Way Businesses Are Legally Structured?**

The tax treatment of income derived from business activity will differ in significant ways depending on the legal form of the business entity. Many large corporations pay the corporate income tax on their profits. But “pass-through businesses” in which profits are “passed through” to owners—including sole proprietorships, partnerships, and S corporations—are reported and taxed under the individual income tax. In that case, the resulting tax liability depends on the total amount and composition of an owner’s income sources, as well as certain demographic characteristics and details related to the tax filing. The Congressional Budget Office generally allocates profits of pass-through businesses to each individual owner, so the choice of a business’s legal structure also matters for the agency’s distributional analysis.

In general, baseline projections from the Congressional Budget Office reflect scheduled tax changes under current law. At the end of 2025, certain provisions of the 2017 tax act are scheduled to expire, which will increase the tax rate on income earned within pass-through entities—whereas the corporate income tax rate of 21 percent will remain unchanged. Thus, even just projecting what will happen based on current law requires estimates of how changes in tax law affect the legal structure of businesses. Much of the existing literature on that front focuses on periods before the increase in pass-through business activity that occurred in the past few decades (Goolsbee 1998; Mackie-Mason and Gordon 1997), or on areas outside the United States (Tazhitdinova 2020).

The role of taxes in determining the legal structure of businesses has distributional implications, too. For example, tax-motivated changes in organizational structure over time account for about one-third of the decline in the measured share of income going to labor (Smith et al. 2022), because income from running a pass-through business is treated as “capital income,” rather than labor income. That effect on measurement arises in part because of differences in the taxation of

wages and business income between corporations and other pass-through entities (CBO 2012).

## Conclusion

The Congressional Budget Office gathers information from the research community in many ways, including consulting the research literature, meeting with experts, obtaining written feedback on draft material, and discussing its work in seminars and conferences. Researchers sometimes work temporarily at CBO for four to twelve months as dissertation fellows or visiting scholars. They may collaborate on research published in professional journals or disseminated to staff on Capitol Hill or to the general public, using many sources of data including those assembled by CBO for policy analysis. Such researchers may also analyze how people would respond to legislative proposals and help develop models that incorporate those responses. In addition, CBO has two panels of advisers: one focused on health and the other on the macroeconomy. Their members represent a variety of perspectives, enabling the agency to gather information and insights from experts with diverse views as well as from the interactions between those experts at panel meetings. CBO's work benefits greatly from its engagement with the research community and especially from the willingness of researchers to spend time providing input that is tailored to help answer questions posed by the US Congress.

## References

- Aerts, Jeroen C. J. H., W. J. Wouter Botzen, Kerry Emanuel, Ning Lin, Hans de Moel, and Erwann O. Michel-Kerjan. 2014. "Evaluating Flood Resilience Strategies for Coastal Megacities." *Science* 344 (6183): 473–75.
- Aleksynska, Mariya, and Ahmed Tritah. 2015. "The Heterogeneity of Immigrants, Host Countries' Income and Productivity: A Channel Accounting Approach." *Economic Inquiry* 53 (1): 150–72.
- Arnold, Robert W. 2018. "How CBO Produces Its 10-Year Economic Forecast." Congressional Budget Office Working Paper 2018-02.
- Berger, Mark C., and Dan A. Black. 1992. "Child Care Subsidies, Quality of Care, and the Labor Supply of Low-Income, Single Mothers." *Review of Economics and Statistics* 74 (4): 635–42.
- Black, Sandra E., Lesley J. Turner, and Jeffrey T. Denning. 2023. "PLUS or Minus? The Effect of Graduate School Loans on Access, Attainment, and Prices." NBER Working Paper 31291.
- Bom, Pedro R. D., and Jenny E. Ligthart. 2014. "What Have We Learned from Three Decades of Research on the Productivity of Public Capital?" *Journal of Economic Surveys* 28 (5): 889–916.
- Borowsky, Jonathan, Jessica H. Brown, Elizabeth E. Davis, Chloe Gibbs, Chris M. Herbst, Aaron Sojourner, Erdal Tekin and Matthew J. Wiswall. 2022. "An Equilibrium Model of the Impact of Increased Public Investment in Early Childhood Education." NBER Working Paper 30140.

- Bradt, Jacob T., and Joseph E. Aldy.** 2023. "Private Benefits from Public Investment in Climate Adaptation and Resilience." Unpublished.
- Carloni, Dorian.** 2021. "Revisiting the Extent to Which Payroll Taxes Are Passed through to Employees." CBO Working Paper 2021-06.
- Cascio, Elizabeth U., and Diane Whitmore Schanzenbach.** 2013. "The Impact of Expanding Access to High-Quality Preschool." *Brookings Papers on Economic Activity* 44 (2): 127–92.
- Clemens, Jeffrey, and Joshua D. Gottlieb.** 2014. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *American Economic Review* 104 (4): 1320–49.
- Coffey, Bentley, Patrick A. McLaughlin, and Pietro Peretto.** 2020. "The Cumulative Cost of Regulations." *Review of Economic Dynamics* 38: 1–21.
- Congressional Budget Office (CBO).** 2007. *Potential Cost Savings from the Pre-Disaster Mitigation Program*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2012. *The Taxation of Capital and Labor Through the Self-Employment Tax*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2013. *The Economic Impact of S. 744, the Border Security, Economic Opportunity, and Immigration Modernization Act*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2014. *How CBO Analyzes the Effects of Changes in Federal Fiscal Policies on the Economy*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2016. *Potential Increases in Hurricane Damage in the United States: Implications for the Federal Budget*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2018a. *An Overview of CBO's Microsimulation Tax Model*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2018b. *H.R. 6106, Common Sense Permitting Act*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2019a. *An Overview of CBO's Life-Cycle Growth Model*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2019b. *S. 1976, Federal Permitting Reform and Jobs Act*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2019c. *Expected Costs of Damage from Hurricane Winds and Storm-Related Flooding*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2020a. *How CBO and JCT Analyze Major Proposals That Would Affect Health Insurance Choices*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2020b. *The Effects of Pandemic-Related Legislation on Output*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2021a. *Budgetary Effects of Climate Change and of Potential Legislative Responses to It*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2021b. *CBO's Policy Growth Model*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2021c. *Economic Effects of Expanding Home- and Community-Based Services in Medicaid*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2021d. *Economic Effects of Expanding Subsidized Child Care and Providing Universal Preschool*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2021e. *Economic Effects of Offering a Federal Paid Family and Medical Leave Program*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2021f. *Effects of Physical Infrastructure Spending on the Economy and the Budget Under Two Illustrative Scenarios*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2021g. *HISIM2: The Health Insurance Simulation Model Used in Preparing CBO's July 2021 Baseline Budget Projections*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2021h. *The Budgetary Effects of the Raise the Wage Act of 2021*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2022a. *Army Corps of Engineers: Budgetary History and Projections*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2022b. *Budgetary Effects of a Policy That Would Lower the Age of Eligibility for Medicare to 60*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2022c. *FEMA's Disaster Relief Fund: Budgetary History and Projections*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2022d. *How CBO Analyzes the Effects of Charging the Oil and Gas Industry for Methane Emissions*. Washington, DC: CBO.

- Congressional Budget Office (CBO).** 2022e. *Options for Reducing the Deficit, 2023 to 2032—Volume I: Larger Reductions*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2022f. *Policy Approaches to Reduce What Commercial Insurers Pay for Hospitals' and Physicians' Services*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2023a. *An Analysis of the Navy's Fiscal Year 2024 Shipbuilding Plan*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2023b. *An Introduction to the Congressional Budget Office*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2023c. *CBO Describes Its Cost-Estimating Process*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2023d. *CBO Explains How It Develops the Budget Baseline*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2023e. *Costs of the Proposed Income-Driven Repayment Plan for Student Loans*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2023f. *Estimated Budgetary Effects of Divisions A and B of H.R. 1, the Lower Energy Costs Act*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2023g. *Estimates of the Cost of Federal Credit Programs in 2024*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2023h. *Flood Damage and Federally Backed Mortgages in a Changing Climate*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2023i. *How CBO Projects Corporate Income Tax Revenues*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2023j. *Medicaid Baseline Projection*. Washington, DC: CBO. [www.cbo.gov/system/files/2023-05/51301-2023-05-medicaid.pdf](http://www.cbo.gov/system/files/2023-05/51301-2023-05-medicaid.pdf).
- Congressional Budget Office (CBO).** 2023k. *The Budgetary and Economic Effects of S. 2488, the Raise the Wage Act of 2023*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2023l. *The Distribution of Household Income, 2020*. Washington, DC: CBO.
- Congressional Budget Office (CBO).** 2024. *The Budget and Economic Outlook: 2024 to 2034*. Box 2-1. Washington, DC: CBO.
- Congressional Research Service.** 2022. *Judicial Review and the National Environmental Policy Act of 1969*. Washington, DC: CBO.
- Council on Environmental Quality.** 2021. *A Citizen's Guide to NEPA*. Washington, DC: Executive Office of the President.
- Curto, Vilsa, Liran Einav, Amy Finkelstein, Jonathan Levin, and Jay Bhattacharya.** 2019. "Health Care Spending and Utilization in Public and Private Medicare." *American Economic Journal: Applied Economics* 11 (2): 302–32.
- Dafny, Leemore S.** 2005. "How Do Hospitals Respond to Price Changes?" *American Economic Review* 95 (5): 1525–47.
- Davlasheridze, Meri and Qing Miao.** 2021. "Does Post-Disaster Aid Promote Community Resilience? Evidence from Federal Disaster Programs." *Natural Hazard* 109: 63–88.
- Dawson, John W., and John J. Seater.** 2013. "Federal Regulation and Aggregate Economic Growth." *Journal of Economic Growth* 18: 137–77.
- Djankov, Simeon, Caralee McLiesh, and Rita Maria Ramalho.** 2006. "Regulation and Growth." *Economic Letters* 92 (3): 395–401.
- Dourado, Eli, and Josh T. Smith.** 2020. *Update to the Regulation Implementing the Procedural Provisions of the National Environmental Policy Act*. Logan, UT: Center for Growth and Opportunity, Utah State University.
- Dunn, Abe, Joshua D. Gottlieb, Adam Shapiro, Daniel J. Sonnenstuhl, and Pietro Tebaldi.** 2021. "A Denial a Day Keeps the Doctor Away." NBER Working Paper 29010.
- Fernald, John, and Huiyu Li.** 2022. "The Impact of COVID on Productivity and Potential Output." Federal Reserve Bank of San Francisco Working Paper 2022-19.
- Finkelstein, Amy, Nathaniel Hendren, and Erzo F. P. Luttmer.** 2019. "The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment." *Journal of Political Economy* 127 (6): 2836–74.
- Fitzpatrick, Maria Donovan.** 2010. "Preschoolers Enrolled and Mothers at Work? The Effects of Universal Prekindergarten." *Journal of Labor Economics* 28 (1): 51–85.

- Gamber, Edward.** 2020. *The Historical Decline in Real Interest Rates and Its Implications for CBO's Projections*. CBO Working Paper 2020-09.
- Goda, Gopi Shah, Ezra Golberstein, and David C. Grabowski.** 2011. "Income and the Utilization of Long-Term Care Services: Evidence from the Social Security Benefit Notch." *Journal of Health Economics* 30 (4): 719–29.
- Goolsbee, Austan.** 1998. "Taxes, Organizational Form, and the Deadweight Loss of the Corporate Income Tax." *Journal of Public Economics* 69 (1): 143–52.
- Government Accountability Office (GAO).** 2023. *Child Care: Subsidy Eligibility and Use in Fiscal Year 2019 and State Program Changes During the Pandemic*. Washington, DC: Government Accountability Office.
- Grabowski, David C., and Jonathan Gruber.** 2007. "Moral Hazard in Nursing Home Use." *Journal of Health Economics* 26 (3): 560–77.
- Gravelle, Jennifer C.** 2010. "Corporate Tax Incidence: Review of General Equilibrium Estimates and Analysis." CBO Working Paper 2010-03.
- Gravelle, Jennifer C.** 2011. "Corporate Tax Incidence: A Review of Empirical Estimates and Analysis." CBO Working Paper 2011-01.
- Greenstone, Michael.** 2002. "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufacturers." *Journal of Political Economy* 110 (6): 1175–1219.
- Habib, Bilal, and Rebecca Heller.** 2022. "Current Work on the Distributional Analysis of Household Income Resulting from Policy Changes." CBO Working Paper 2022-09.
- Johnson, Kris A., Oliver E. Wing, Paul D. Bates, Joseph Fargione, Timm Kroeger, William D. Larson, Christopher C. Sampson, and Andrew M. Smith.** 2019. "A Benefit–Cost Analysis of Floodplain Land Acquisition for US Flood Damage Reduction." *Nature Sustainability* 3: 56–62.
- Karamcheva, Nadia, Jeffrey Perry, and Constantine Yannelis.** 2020. "Income-Driven Repayment Plans for Student Loans." CBO Working Paper 2020-02.
- Kargar, Mahyar, and William Mann.** 2023. "The Incidence of Student Loan Subsidies: Evidence from the PLUS Program." *Review of Financial Studies* 36 (4): 1621–66.
- Kelchen, Robert.** 2019. "An Empirical Examination of the Bennett Hypothesis in Law School Prices." *Economics of Education Review* 73: Article 101915.
- Kelly, David L., and Renato Molina.** 2023. "Adaptation Infrastructure and its Effects on Property Values in the Face of Climate Risk." *Journal of the Association of Environmental and Resource Economists* 10 (6): 1405–38.
- Kiska, Wendy, Jason Levine, and Damien Moore.** 2017. "Modeling the Costs of the Pension Benefit Guaranty Corporations' Multiemployer Program." CBO Working Paper 2017-04.
- Lasky, Mark.** 2022. "The Congressional Budget Office's Small-Scale Policy Model." CBO Working Paper 2022-08.
- Lewis, Eric.** 2019. "Patchwork Policies, Spillovers, and the Search for Oil and Gas." *American Economic Journal: Economic Policy* 11 (1): 380–405.
- Li, Hao.** 2020. "The Effect of Universal Pre-kindergarten Policy on Female Labor Force Participation—A Synthetic Control Approach." *Southern Economic Journal* 87 (2): 440–82.
- Love, David A., Paul A. Smith, and David W. Wilcox.** 2011. "The Effect of Regulation on Optimal Corporate Pension Risk." *Journal of Financial Economics* 101 (1): 18–35.
- Lucca, David O., Taylor Nadauld, and Karen Shen.** 2019. "Credit Supply and the Rise in College Tuition: Evidence from the Expansion in Federal Student Aid Programs." *Review of Financial Studies* 32 (2): 423–66.
- Mackie-Mason, Jeffrey K., and Roger Gordon.** 1997. "How Much Do Taxes Discourage Incorporation?" *Journal of Finance* 52 (2): 477–505.
- Michalopoulos, Charles.** 2010. *Effects of Reducing Child Care Subsidy Copayments in Washington State*. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, US Department of Health and Human Services.
- Multihazard Mitigation Council.** 2005. *Natural Hazard Mitigation Saves: An Independent Study to Assess the Future Savings from Mitigation Activities*. Washington, DC: National Institute of Building Sciences.
- National Institute of Building Sciences.** 2019. *Natural Hazard Mitigation Saves: 2019 Report*. Washington, DC: National Institute of Building Sciences.
- Neumann, James E., Paul Chinowsky, Jacob Helman, Margaret Black, Charles Fant, Kenneth Strzepek, and Jeremy Martinich.** 2021. "Climate Effects on US Infrastructure: The Economics of Adaptation for Rail, Roads, and Coastal Development." *Climatic Change* 167: Article 44.



- Ortega, Francesc, and Giovanni Peri.** 2009. "The Causes and Effects of International Migrations: Evidence from OECD Countries 1980–2005." NBER Working Paper 14833.
- Ortega, Francesc, and Giovanni Peri.** 2014a. "Openness and Income: The Roles of Trade and Migration." *Journal of International Economics* 92 (2): 231–51.
- Ortega, Francesc, and Giovanni Peri.** 2014b. "The Aggregate Effects of Trade and Migration: Evidence from OECD Countries." In *The Socio-Economic Impact of Migration Flows: Effects on Trade, Remittances, Output, and the Labour Market*, edited by Andrés Artal-Tur, Giovanni Peri, Francisco Requena-Silvente, 19–51. Berlin: Springer International Publishing.
- Patterson, Christina, Jonathan Petkun, and William Skimmyhorn.** 2019. "I Want You! (But Not You): Selection in Military Retention." Unpublished.
- Peri, Giovanni.** 2012. "The Effect of Immigration on Productivity: Evidence from U.S. States." *Review of Economics and Statistics* 94 (1): 348–58.
- Perry, Mark J.** 2022. "Chart of the Day...or Century?" American Enterprise Institute, July 23, [www.aei.org/carpe-diem/chart-of-the-day-or-century-8](http://www.aei.org/carpe-diem/chart-of-the-day-or-century-8).
- Philippon, Thomas.** 2022. "Additive Growth." NBER Working Paper 29950.
- Prato, Marta.** 2022. "The Global Race for Talent: Brain Drain, Knowledge Transfer, and Growth." <http://dx.doi.org/10.2139/ssrn.4287268>.
- Rachel, Lukasz, and Thomas D. Smith.** 2017. "Are Low Real Interest Rates Here to Stay?" *International Journal of Central Banking* 13 (3): 1–42.
- Rauh, Joshua D., Irina Stefanescu, and Stephen P. Zeldes.** 2020. "Cost Saving and the Freezing of Corporate Pension Plans." *Journal of Public Economics* 188: 104211.
- Reichling, Felix, and Shinichi Nishiyama.** 2015. "The Costs to Differnet Generations of Policies That Close the Fiscal Gap." CBO Working Paper 2015-10.
- Ryan, Stephen P.** 2012. "The Costs of Environmental Regulation in a Concentrated Industry." *Econometrica* 80 (3): 1019–61.
- Saez, Emmanuel, and Gabriel Zucman.** 2023. *Distributional Tax Analysis in Theory and Practice: Harberger Meets Diamond-Mirrlees*. NBER Working Paper 31912.
- Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick.** 2022. "The Rise of Pass-Throughs and the Decline of the Labor Share." *American Economic Review: Insights* 4 (3): 323–40.
- Tazhitdinova, Alisa.** 2020. "Are Changes in Organizational Form Costly? Income Shifting and Business Entry Responses to Taxes." *Journal of Public Economics* 186: 104187.
- Van Houtven, Courtney Harold, Brian E. McGarry, Eric Jutkowitz, and David C. Grabowski.** 2020. "Association of Medicaid Expansion under the Patient Protection and Affordable Care Act with Use of Long-term Care." *JAMA Network Open* 3 (10): Article e2018728.
- Wobus, Cameron, Jeremy Porter, Mark Lorie, Jeremy Martinich, and Rachel Bash.** 2021. "Climate Change, Riverine Flood Risk and Adaptation for the Conterminous United States." *Environmental Research Letters* 16 (9): 094034.
- Wu, Vivian Y., and Yu-Chu Shen.** 2014. "Long-Term Impact of Medicare Payment Reductions on Patient Outcomes." *Health Services Research* 49 (5): 1596–1615.
- Yannelis, Constantine, and Greg Tracey.** 2022. "Student Loans and Borrower Outcomes." *Annual Review of Financial Economics* 14: 167–86.XX



# The Economic Constitution of the United States

Cass R. Sunstein

**S**uppose that the US Department of Transportation wants to issue a new regulation, one that would require all new motor vehicles in the United States to be equipped with some state-of-the-art safety technology. Will the regulation go forward? The answer might well lie in the hands of the Office of Information and Regulatory Affairs (OIRA), a small office in the Office of Management and Budget. Headed by an administrator who is nominated by the president and confirmed by the Senate, OIRA consists of policy analysts, economists, and lawyers, who will ask whether the benefits of the new regulation would justify the costs. That question is asked not only for motor vehicle safety regulations; it is also asked for climate change regulations, occupational safety regulations, water pollution regulations, immigration regulations, animal welfare regulations, airline safety regulations, and many others as well. But how do agencies assess costs and benefits? Where will they look?

As it turns out, the United States has something like an Economic Constitution, designed to answer those questions. Most Americans, and even most economists, know nothing about it. Its focus is on human welfare, understood essentially in the economic terms of cost-benefit analysis, with occasional doses of political philosophy (emphasizing, for example, obligations to future generations and the

■ *Cass R. Sunstein is the Robert Walmsley University Professor, Harvard Law School, Cambridge, Massachusetts. From 2009 to 2012, Sunstein served as Administrator, Office of Information and Regulatory Affairs, Washington, DC. From 2021 to the present, he has also served as senior counselor and consultant with the Department of Homeland Security, where he has occasionally been involved in discussions of some of the issues explored in this essay. His email address is [csunstei@law.harvard.edu](mailto:csunstei@law.harvard.edu).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.2.25>.

difficulty of monetizing human dignity). It was developed by the Office of Management and Budget and the Council of Economic Advisers, and after public comment and peer review, was originally issued on September 17, 2003, under the name of OMB Circular A-4 (OMB 2003). The basic goal of Circular A-4 was to systemize existing learning on how to monetize costs and benefits and on what to do in the face of uncertainty. The circular was intended to provide guidance for disparate agencies, which might otherwise be at sea, and might be working on the basis of inconsistent assumptions or make serious economic mistakes. Circular A-4 applies to an extraordinary range of regulatory topics: fuel economy standards, silica in the workplace, greenhouse gas emissions, rearview cameras in automobiles, food safety requirements, and many others (Sunstein 2018).

The requirement to demonstrate that the benefits of regulation justify the costs raises fundamental questions. When is federal regulation a good idea in the first instance (Breyer 1980)? How should agencies assign monetary values to mortality and morbidity reductions (Viscusi 1993)? What is the right discount rate for effects that will occur in a decade or in many decades (Weitzman 1998)? How, if at all, does distribution matter (Adler 2016)? How should agencies handle uncertainty, as in cases in which the goal is to reduce the risk of a financial crisis or a terrorist attack (Sunstein 2021a)? The Economic Constitution seeks to summarize the best available economic thinking on these and other questions.

To understand the origins of the Economic Constitution, we need to back up a bit. The Office of Information and Regulatory Affairs was created by the Paperwork Reduction Act, signed into law by President Jimmy Carter in 1980. The Paperwork Reduction Act responded to a specific problem: the imposition of excessive paperwork burdens, seen as a potential obstacle to economic activity and likely to impose particular harm on the most vulnerable members of society (including people who are elderly or in poor health). OIRA's authority was greatly expanded by President Ronald Reagan, who saw it not only as a check on paperwork burdens, but also and far more broadly as a brake on unjustified regulation and as a repository of expertise with respect to economic analysis. Executive Order 12291, written by both economists and lawyers, was signed by Reagan in 1981. It established OIRA's role in overseeing national regulation. The fundamentals of that Executive Order continue in effect to this day. It set out five general requirements:

1. Administrative decisions shall be based on adequate information concerning the need for and consequences of proposed government action;
2. Regulatory action shall not be undertaken unless the potential benefits to society for the regulation outweigh the potential costs to society;
3. Regulatory objectives shall be chosen to maximize the net benefits to society;
4. Among alternative approaches to any given regulatory objective, the alternative involving the least net cost to society shall be chosen; and
5. Agencies shall set regulatory priorities with the aim of maximizing the aggregate net benefits to society, taking into account the condition of the

particular industries affected by regulations, the condition of the national economy, and other regulatory actions contemplated for the future.

To implement these requirements, Executive Order 12291 required all executive agencies of the federal government to produce a Regulatory Impact Analysis of major regulations. Among other things, the Regulatory Impact Analysis must include a description of benefits and costs, whether quantifiable or not; a description of the net benefits (or net costs) of the regulation in question; and a description of alternative approaches that could achieve the same regulatory goal at lower cost. Remarkably, these directives have been explicitly ratified or followed by all subsequent presidents, most recently by President Joe Biden (Executive Order 14094 [2023]).

Within the executive branch, the Reagan-Bush-Clinton-Bush-Obama-Trump-Biden consensus (!), as we might call it, rests on a theory: Executive agencies should promote human welfare, and an accounting of costs and benefits is one way to make it more likely that they will do that (Adler and Posner 2006; Sunstein 2018; Sen 2000). The idea of “welfare” can of course be understood in many ways; within the government, the operating consensus has been that if the monetized costs of regulation exceed the monetized benefits, the effects on welfare are not likely to be good (Adler and Posner 2006). To be sure, Clinton also directed agencies to consider “distributive impacts” and “equity” (Executive Order 12866 [1993]), and Obama underlined those directives and also added a reference to “human dignity” (Executive Order 13563 [2011]). In Executive Order 13771 (2017), Trump imposed a “one in, two out” approach to regulation, by which agencies had to eliminate two regulations for every regulation that they issued.

But all the while, Circular A-4, with its focus on cost-benefit analysis as a proxy for human welfare, emerged unscathed. Indeed, it proved remarkably enduring; it stabilized and oriented the development of regulatory analysis for two decades. Until 2023, no president changed a word of it. It follows that when agencies submitted significant rules to the Office of Information and Regulatory Affairs, they were required to accompany those rules with a Regulatory Impact Analysis, which abided by the circular. Agencies made their best efforts to do so, but before rules and accompanying analyses were made public, there were often extended discussions with OIRA about whether they had done an adequate job. Some of those discussions might occur before agencies submit their analyses to OIRA; this might be true for especially challenging issues (consider how to place a value on prevention of prison rape). Economists in the Executive Office of the President (including the Council of Economic Advisers and the National Economic Council) might well have become involved. There was often considerable back-and-forth on technical issues, involving multiple rounds of discussions, with OIRA having the final word on whether the economic analysis was adequate. Members of the public (including economists) might have been involved as well, both through the public comment process, required by law, and through requests for meetings (“12866 meetings,” as

they were called). Before rules and analyses were produced, officials might have engaged with academic economists on challenging questions (though this was rare).

On his first day in office, President Biden (2021) took the extraordinary step of explicitly directing the Office of Management and Budget to provide recommendations for “modernizing” the process of regulatory review, prominently through revisions to Circular A-4. Biden specified that the revisions should (1) “reflect new developments in scientific and economic understanding,” (2) account for “regulatory benefits that are difficult or impossible to quantify,” and (3) ensure against “harmful anti-regulatory or deregulatory efforts.” In particular, he emphasized the importance of taking “into account the distributional consequences of regulation.”

In early 2023, the Biden administration put forward its long-awaited proposed revision to Circular A-4 for public comment and for peer review (OMB 2023a; OMB 2023b). The proposed revision was much longer and more detailed than the original. It did maintain fundamental continuity with the 2003 document, and in some ways, that may be the biggest news of all. Importantly, the Office of Information and Regulatory Affairs process, as sketched above, was not affected at all; it continues in the same form. But drawing on the last two decades of economic research, the proposed revision also suggested significant changes (see the lengthy preamble, OMB 2023a), and offered elaboration in places where the 2003 document was brief and in some ways cryptic. In late 2023, and after an extensive process of public comment and peer review, the Biden administration finalized a new Circular A-4 (OMB 2023c). The Economic Constitution of the United States was reborn.

## **Justifications for Regulation: Market Failures and Behavioral Economics**

When should the government regulate at all? The original Circular A-4 offers an account of why regulation might be justified and directs federal agencies to accompany their analyses with what they take to be the proper justification. It was evidently influenced by canonical texts (Breyer 1980), in that it gives pride of place to standard economic accounts: externalities, common property resources, public goods, market power, and inadequate or asymmetric information. Interestingly, however, the original Circular A-4 also built on early work in behavioral economics (Kahneman, Slovic, and Tversky 1982; Thaler 1994) to explain the relevance of bounded rationality:

Even when adequate information is available, people can make mistakes by processing it poorly. Poor information-processing often occurs in cases of low probability, high-consequence events, but it is not limited to such situations. For instance, people sometimes rely on mental rules-of-thumb that produce errors. If they have a clear mental image of an incident which makes it cognitively available, they might overstate the probability that it will occur. Individuals sometimes process information in a biased manner, by being too

optimistic or pessimistic, without taking sufficient account of the fact that the outcome is exceedingly unlikely to occur.

The original Circular A-4 also briefly recognizes certain justifications for regulation that have nothing to do with conventional market failures, including an increase in the efficiency of government, redistribution, protection of privacy, increasing freedom, or promoting “other democratic aspirations.”

The new and revised version of Circular A-4 embraces these ideas, but with close reference to more recent economic research, it greatly expands on them. Importantly, it too gives pride of place to the standard market failures, while noting that externalities “can also be associated with positional goods, which can exist if any increase in the relative position of one person lowers the relative position of others (and vice versa)” (Frank 2005). The new circular emphasizes that some regulations might be designed to increase government efficiency and improve public services; consider, for example, the TSA Precheck program that allows travelers to register in advance to move through airport security more quickly, the Global Entry program that facilitates entry into the United States for low-risk travelers, the rise of mobile drivers’ licenses that can be displayed on a smartphone, and more generally the reduction of administrative burdens or “sludge” (Sunstein 2021b). The new circular notes that regulation might protect civil rights and civil liberties and advance democratic values. It adds that regulation might promote distributional fairness and advance equity (more on this below).

Going far beyond the brief treatment in 2003 (and this shows a great deal of learning), the new Circular A-4 also offers an extensive discussion of “behavioral biases.” It discusses “internalities”: “When individuals exhibit imperfect self-control, they make a decision that increases short-term well-being by less than it decreases future well-being (appropriately discounted).” It refers specifically to a number of behavioral biases: availability bias, stemming from the availability heuristic (people often evaluate risks by asking whether a salient example, such as a workplace accident, comes readily to mind and so is cognitively “available”); present bias (people might care much more about the short-term than the long-term, and might apply an indefensibly high discount rate to future benefits or costs); unrealistic optimism (people might think that things are more likely to turn out well than a realistic appraisal suggests); and status quo bias (leading to a preference for inaction) (Samuelson and Zeckhauser 1988). The new Circular A-4 explicitly directs agencies to “consider the degree to which the evidence available to you indicates that behavior reflects rational preferences and the degree to which it indicates that such behavior is the product of a behavioral bias.”

In drawing attention to behavioral biases, the discussion in the new circular builds on a great deal of earlier work in behavioral economics; for example, it cites Tversky and Kahneman (1973), Chetty (2015), Thaler (2016), and Gilovich, Griffin, and Kahneman (2002). The analysis is also consistent with the relevant analyses in Gabaix (2019), which suggests that inattention might be an organizing principle for diverse behavioral findings, and in Rabin (2013), which explores paternalism

and the formation or nonformation of healthy habits. We might even see Circular A-4 itself as an attempted corrective to behavioral biases on the part of regulators themselves—as, for example, where availability bias leads officials to exaggerate some risks and downplay others (Sunstein 2018).

Distinguishing among the variety of regulatory tools, the new circular explicitly notes that the best response to behavioral biases might involve information disclosure and “nudges”:

Measures that serve as nudges—such as changing the default or pre-selected options, or changing the manner in which information is presented—can also improve consumer welfare without restricting choice. Such nudges can include simplifying choices through sensible default rules (such as setting automatic enrollment with opt-out versus opt-in); reducing complexity; requiring active choice; increasing the salience of certain factors or variables; and promoting desirable social norms.

Indeed, the new circular specifies that “nudges make most sense when the market failure involves a behavioral bias, although even in such cases, nudges may not be either appropriate or sufficient.” It adds that “informational measures or nudges, like other measures, should be evaluated in terms of their benefits and costs.” This suggestion, emphasizing the need for a careful analysis of the welfare effects of responses to behavioral biases, is a central theme in the research literature (among others, Viscusi 2022; Thunström 2019; Rabin 2013; Allcott and Kessler 2019). A vexing area that deserves further research is the effect of disclosure requirements. People might change their behavior (by, for example, making healthier food choices), but they might not enjoy their meals as much, and they might not enjoy being nudged; those welfare losses should be counted (Thunström 2019; Sunstein 2020). At the same time, preferences might turn out to be shifting and endogenous to the nudge (consider the potential development of tastes for healthier foods), which raises further complications.

## **Valuing Life**

In the domain of health and safety, the monetized benefits of regulations often come largely from reductions in mortality risks, and the original version of Circular A-4 offers a detailed discussion of how to monetize those reductions. With close reference to the economic literature (Viscusi 1993; Viscusi and Aldy 2003), it declares plainly: “The willingness-to-pay approach is the best methodology to use if reductions in fatality risk are monetized.” Hence the document calls for use of the “value of statistical life.” The basic idea is to infer willingness-to-pay from the actual choices people make; say, to what extent do workers receive additional pay for jobs with higher mortality risk, or to what extent are consumers willing to pay for products that reduce their personal risk. As a back-of-the-envelope illustration,



imagine that a job with a mortality risk that involves one additional death for every 10,000 workers pays \$1,000 more per year. The value of a statistical life would then be \$10 million—that is,  $\$1000 \times 10,000$ .

The original Circular A-4 notes that “valuation of fatality risk reduction is an evolving area in both results and methodology,” and that “literature-based VSL estimates may not be entirely appropriate for the particular risk being evaluated (e.g., the use of occupational risk premia to value reductions in risks from environmental hazards),” which means that agencies should explain “any adjustments of the estimates to reflect the nature of the risk” that is at stake (for discussion, see Viscusi 2010).

The original Circular A-4 also draws the attention of regulatory agencies to the disputed question (Sunstein 2004) whether valuation should vary depending on the context and the affected population—for example, by using the value of a statistical life-year (VSLY), which implies that saving the life of a younger person, with more expected years of life, is worth more than saving the life of an elderly person. Thus, the document notes that some people emphasize “that the value of a statistical life is not a single number relevant for all situations,” and that agencies “should consider providing estimates of both VSL and VSLY, while recognizing the developing state of knowledge in this area.”

At the same time, and importantly, the original Circular A-4 notes that the existing literature generally suggests a VSL of between \$1 million and \$10 million. It also points to the continuing “special challenges” in valuing children’s lives, in light of the fact that agencies cannot rely on a child’s willingness to pay for health improvements, and a parent’s willingness to pay, to reduce a risk faced by a child, “may need to be expanded to include a societal interest in child health and safety.” Some work does attempt to estimate how much parents would pay to reduce risks to their children (Williams 2013), but even if we could rely on those estimates, they do not include the welfare of children themselves, just as people’s willingness to pay to reduce risks to animals does not include the welfare of animals themselves (Carlson et al. 2019). Valuation of the benefits of reducing risks to children remains a singularly vexing issue, on which far more work remains to be done; current understandings are limited (Robinson et al. 2019).

Interestingly, the new Circular A-4 changes almost nothing in the 2003 text. It accepts the theoretical analysis, with all of the relevant cautionary notes. The preamble to the proposed version announces this choice without offering a great deal of explanation: “While recognizing that potential modifications to material on monetizing health and safety benefits and costs and health and safety metrics could be advantageous, OMB believes that continued reliance on this material is generally appropriate at this time. . . . OMB does not intend to substantially revise this material at this time” (OMB 2023a). There is one exception, and it is important. The new circular notes that “agencies utilize central estimates of VSL between \$10 million to \$12 million as of 2022, and regularly update these values to reflect inflation and real income growth.”

It is worthwhile to ask why the new circular did not rethink the earlier treatment in a more fundamental way. One possibility is a judgment that in the last two

decades, economic understanding of how to value statistical lives has not significantly changed, even if there have been many new studies and methodological improvements (in this journal, Lavetti 2023).

As the new circular notes, there are also difficulties in monetizing the benefits that come from reducing morbidity rather than mortality. Here is the text: “Suppose further that baseline evidence indicates 100,000 individuals experience non-fatal health harms—perhaps such that they stay home from work or school—on an average of two days per year, due to the pollution. Two challenging areas for estimating morbidity-related regulatory benefits might be quantifying the regulation’s effectiveness at reducing health harm and monetizing the per-day benefit of avoiding the health harm.” In principle, one might think that a standard approach is the right one: How much do people pay to eliminate a  $1/x$  risk of (say) a nonfatal heart attack, or how much would they demand to face such a risk? But there is not a great deal of evidence on questions of this kind.

## Discount Rates

Suppose that a regulation would impose \$800 million in costs, incurred largely over the next two years, but deliver \$950 million in benefits, to be enjoyed over a period starting ten years from now. Or suppose that a regulation would impose \$2 billion in costs, to be incurred mostly over the next five years, but deliver \$4 billion in benefits, to be enjoyed mostly by future generations. How should future benefits be valued? What is the appropriate discount rate? On this question, the original Circular A-4 offers three points (OMB 2003):

1. Resources that are invested will normally earn a positive return, so current consumption is more expensive than future consumption, because you are giving up that expected return on investment when you consume today.
2. Postponed benefits also have a cost because people generally prefer present to future consumption. They are said to have positive time preference.
3. Also, if consumption continues to increase over time, as it has for most of US history, an increment of consumption will be less valuable in the future than it would be today, because the principle of diminishing marginal utility implies that as total consumption increases, the value of a marginal unit of consumption tends to decline.

This general account is broadly in line with some standard accounts in the economic literature (Arrow et al. 2014). But somewhat confusingly, the original Circular A-4 calls for use of two discount rates: 7 percent and 3 percent. The higher figure is meant to capture “the average before-tax rate of return to private capital in the US economy,” and “reflects the returns to real estate and small business capital as well as corporate capital.” Under the earlier version of Circular A-4, the 7 percent is stated to be the “base-case for regulatory analysis.” The lower figure is meant to

recognize that the “effects of regulation do not always fall exclusively or primarily on the allocation of capital.” It follows that if “regulation primarily and directly affects private consumption (for example, through higher consumer prices for goods and services), a lower discount rate is appropriate.” The 3 percent figure is meant to capture the social rate of time preference. Hence, Circular A-4 stated that agencies “should provide estimates of net benefits using both 3 percent and 7 percent” (OMB 2003).

The intergenerational case presents its own special ethical considerations for choosing a discount rate. The original Circular A-4 noted: “Although most people demonstrate time preference in their own consumption behavior, it may not be appropriate for society to demonstrate a similar preference when deciding between the well-being of current and future generations. Future citizens who are affected by such choices cannot take part in making them, and today’s society must act with some consideration of their interest.” At the same time, the circular insists that “it would still be correct to discount future costs and consumption benefits generally (perhaps at a lower rate than for intragenerational analysis), due to the expectation that future generations will be wealthier and thus will value a marginal dollar of benefits or costs by less than those alive today.”

The original Circular A-4 also drew attention to “increased uncertainty about the appropriate value of the discount rate, the longer the horizon for the analysis.” It added: “As explained by Weitzman (1998), in the limit for the deep future, the properly averaged certainty-equivalent discount factor (i.e.,  $1/[1+r]t$ ) corresponds to the minimum discount rate having any substantial positive probability.” With these points in mind, Circular A-4 authorized agencies to use a lower discount rate, as part of a sensitivity analysis, for rules having significant intergenerational benefits or costs.

The new version of the circular adopts a broadly similar analytic framework, but also a significantly different number, based on the social rate of time preference. It notes that a standard “approach assumes that the real (inflation-adjusted) rate of return on long-term US government debt provides a fair approximation” of that rate. The new circular states that over the last 30 years, the rate has averaged 2.0 percent in real terms, before taxes. It suggests that agencies use that number as that “default rate.” (Interestingly, the proposed revision offered a more detailed discussion of alternatives, including the Ramsey framework; the final version of the new circular does not include that discussion.) A reduction of the discount rate from 7 percent and 3 percent to 2 percent will inevitably have a real effect on projected costs and benefits, perhaps especially in cases in which health and safety benefits (say, from clean air regulations) are expected in the not-immediate future.

This analysis is meant to apply “for all effects from the present through thirty years into the future.” As a practical matter, the vast majority of rules that agencies promulgate, and that OIRA reviews, are modeled as having their principal effects within that period. Whether we are speaking of fuel economy regulations, automobile safety regulations, occupational health regulations, or cybersecurity regulations, a 30-year period (or less) is generally taken to be reasonable.

What about the more distant future? That question is relevant for some regulations, above all those involving climate change. On that question, the new circular is broadly in line with the 2003 document. Referring to a principle of intergenerational neutrality (Cowen and Parfit 1992), it notes (and perhaps can be taken to endorse) the view that “government should treat all generations equally.” For those who do endorse that view, what should be discounted is future money, not future welfare; that is, the welfare of a person born in 2040 is not worth less than that of a person born in 1990. Building on work by Weitzman (1998) and many others (Arrow et al. 2014), the new circular adds: “Because future changes in the social rate of time preference are uncertain but correlated over time, the certainty-equivalent discount rate will have a declining schedule.” The new circular has an appendix with relevant numbers—for example, 1.7 percent from 2106 to 2115 (OMB 2023d). This analysis obviously bears on the problem of climate change, though the new circular does not discuss the social cost of carbon, which is being handled separately (EPA 2023).

It is predictable that there will be continuing theoretical and empirical discussion of whether the 2.0 percent rate is appropriate one; there is a substantial literature on discount rates, some of it raising fundamental questions on the theoretical and empirical levels (Caplin and Leahy 2004; Nordhaus 2007; Weitzman 1994; Heal 2007; Millner and Heal 2023). Is a principle of intergenerational neutrality the right one? (Probably.) But would future generations prefer that we spend money on preventing harms at a 2 percent or 1.7 percent discount rate? Or would they prefer that we sought out higher-return investments (on, say, the education, health, and welfare of children)? Would a somewhat higher, or somewhat lower, discount rate improve the welfare of those who will follow us?

## **Distribution**

Suppose that a regulation would impose \$800 million in annual costs and deliver \$700 million in annual benefits, but that the costs would be imposed mostly on wealthy consumers (who would perhaps have to pay more for a luxury good) and that the benefits would be enjoyed mostly by poor workers (who would have higher income). What then? Do distributional effects matter? Here is an intuitive way to think about it (with a long and contested history in economic thought): If a rich person gets \$10,000, the effect on that person’s welfare will be much less than if a poor person gets \$10,000. If the intuition is correct, then regulators should see a monetary gain to poor people as producing more welfare than an equivalent gain to rich people (Adler 2019).

The original Circular A-4 focused on adding up (appropriate discounted) costs and benefits, and thus offers only a brief discussion of distributional effects. It does ask agencies to “provide a separate description of distributional effects (i.e., how both benefits and costs are distributed among sub-populations of particular concern) so that decision makers can properly consider them along with the effects on economic

efficiency.” It also states that when distributive effects are believed to be important, “the effects of various regulatory alternatives should be described quantitatively to the extent possible, including the magnitude, likelihood, and severity of impacts on particular groups.” There was of course an extensive economic literature on this issue even at the time, with which the circular (somewhat puzzlingly) did not much engage, and which urged that those involved in law and regulation should expand the size of the pie, and leave redistribution to the tax system (Kaplow and Shavell 2002). On that view, regulatory decisions should not be based on distributional issues and should not consider who is helped and who is hurt; instead, maximizing net benefits is the goal. The original Circular A-4 did not embrace that approach, though particular regulatory decisions, under both Democratic and Republican administrations, generally did not engage the distributional questions.

The new circular is continuous with the 2003 document in principle, but it offers much more detail. It emphasizes that an analysis of who is helped and who is hurt might help agencies to “identify alternative regulatory options or impacts that can be mitigated through other regulatory or non-regulatory decisions.” The new circular adds that “it may be useful to analyze the incidence of regulatory effects on each group of interest, or combinations of those groups.” Which groups should be counted? The new circular draws particular attention to “income groups.” It also states that “other economic and demographic categories such as those based on race and ethnicity, sex, gender, geography, wealth, disability, sexual orientation, religion, national origin, age or birth cohort, family composition, or veteran status—among others—may be relevant to a particular regulation.” That is obviously a long list, and the draft gives agencies discretion to say what would be “relevant.”

Importantly, the new circular goes far beyond the original in drawing attention to the possible use of distributional weights, in a way that is meant to account for the diminishing marginal utility of goods as income rises: “Agencies may choose to conduct a benefit-cost analysis that applies weights to the benefits and costs accruing to different groups in order to account for the diminishing marginal utility of goods when aggregating those benefits and costs.” Drawing on a long tradition in economic theory that has rarely played a significant role in regulatory policy in the United States (for relevant discussion, see Adler 2019), the new circular states that because of diminishing marginal utility, “an additional unit of a good is more valuable to a person (in welfare terms) if they have fewer total goods than if they have more total goods.” Agencies have traditionally not used distributional weights in regulatory impact analyses; doing so can be counted as a significant change. There are also risks of legal challenges: Do regulatory statutes allow agencies to consider distributional issues? Is it “arbitrary,” in the legal sense, for them to do so?

There is a great deal of further thinking to do on this question, and a substantial literature on which to draw, some of which could help in the specification of the magnitude of relevant weights (Harberger 1978; Adler 2016; Weisbach 2015; Fleurbaey and Abi-Rafteh 2016; Nurmi and Ahtainen 2018). The problem of incidence deserves particular attention; regulations that seem to benefit particular groups (such as motor vehicle safety regulations) might also impose costs on them (Hemel

2022). Also deserving particular attention (and not mentioned in the new circular) is the recent interest in “prioritarianism,” which emphasizes the importance of focusing on those at the bottom of the welfare ladder (Adler and Norheim 2022). A “prioritarian” social welfare function is different from a utilitarian one, even putting the declining marginal utility of money to one side; it would be useful to engage the potential uses and limits of prioritarianism in the regulatory context.

### **Citizens and Noncitizens, Residents and Nonresidents**

Suppose that a regulation would affect three groups of people: American citizens, residents who are noncitizens, and noncitizens who are nonresidents. Should agencies consider all of those effects? The new Circular A-4 offers a far more detailed discussion than its predecessor. It states that in “many circumstances,” the focus should be on citizens and residents of the United States. At the same time, it notes that if a regulation affects foreign entities, it might affect American citizens and residents as well (for example, by increasing prices). With an implicit nod to the problem of climate change, the new circular notes: “Relevant effects also include the effects of a regulation on US strategic interests, including the potential for inducing strategic reciprocity or other policy changes from actors abroad, or effects on US government assets located abroad. Such effects are particularly likely to occur when your regulation bears on a global commons or a global public good.” If, for example, a regulation reduces greenhouse gas emissions, the beneficial effects on noncitizens outside the United States might prompt strategic reciprocity, in the form of reductions by other nations that ultimately benefit Americans. There might also be effects on Americans living abroad, and those should be taken in account.

The new circular emphasizes the importance of transparency. If the agency is focusing on global effects (perhaps because international or domestic law requires a global calculation), “it is generally appropriate to produce a separate supplementary analysis of the effects experienced by US citizens and residents, unless you determine that such effects cannot be separated in a practical and reasonably accurate manner.” There is room for far more work on these issues, especially on the question whether and when agencies should consider the effects of climate change regulations on the world as a whole (Kotchen 2018).

### **Uncertainty and What Cannot Be Quantified**

One of the most serious challenges for economic analysis of regulations stems from the fact that agencies sometimes know too little to come up with anything like precise numbers (Sunstein 2014). For example, they might not be able to specify the health effects of a reduction in mercury emissions; they might not be able to quantify the benefits from a cybersecurity regulation; they might find it difficult to turn a reduction of prison rapes into monetary equivalents; as noted, they might

struggle to monetize the benefits of a regulation that reduces mortality risks faced mostly by children. With such challenges in mind, the original version of Circular A-4 devotes a great deal of attention to uncertainty. It emphasizes the importance of attempting to offer formal estimates of probabilities: “[Y]our analysis should include two fundamental components: a quantitative analysis characterizing the probabilities of the relevant outcomes and an assignment of economic value to the projected outcomes.”

At the same time, the old Circular A-4 acknowledges the reality of gaps in information, making it impossible to quantify the effects of some regulations. For example, it seems difficult to monetize the effects of disclosure requirements (Thunström 2019) or to monetize the benefits of a rule designed to allow people who use wheelchairs to have access to public buildings. Contingent valuation studies might be used (“how much would you be willing to pay to receive information about fuel economy?”), but they are often unreliable as a guide to welfare effects (Sunstein 2020; Williams 2013). In the hardest cases, Circular A-4 calls for a “threshold” or “break-even analysis,” in which agencies ask: What is the threshold that benefits would have to meet in order for the regulation to have net benefits?

The revised circular is in the same spirit, but it offers significantly more detail. Its central mandate is relatively clear: agencies “should use appropriate statistical techniques to determine a probability distribution of the relevant outcomes.” In some cases, however, it may not be possible to produce such a probability distribution. The new circular emphasizes that nonmonetized benefits and costs might be important, and also pointedly notes: “When it is not possible to monetize all of the important benefits and costs, the alternative with the greatest monetized net benefits will not necessarily be the alternative that generates the greatest social welfare.” In the face of nonmonetized benefits, the new circular again calls for a break-even analysis, which asks what magnitude those benefits “would need to have for the regulation at issue to yield positive net benefits or to change which regulatory alternative is most net beneficial.” Agencies have engaged in break-even analysis on important occasions, as with regulations reducing disability discrimination and targeting prison rape, but much more systematic thinking should be done on that issue, perhaps by directly measuring the subjective experience of those helped and hurt (Sunstein 2018).

There is also a question of whether regulators should assume risk neutrality, risk aversion, or risk-seeking. The new circular says this: “Risk aversion is widespread, and an underlying motivation for insurance and savings behavior.” Speaking modestly, it authorizes agencies to “develop an analysis that takes risk aversion into account.” But when? The modesty of the circular is justified, because there is a great deal more to say and to learn here. If risk aversion is widespread, so is risk-seeking (as emphasized by prospect theory, see Ruggari et al. 2020). If people are risk-averse, they might be thought to be risk-averse with respect to safety and health; but are they, really? And if people are risk-averse, are they also risk-averse with respect to the risks introduced by regulation—as, for example, when fuel economy regulations risk creating less safe cars, or when expensive regulations introduce social risks simply by virtue of

the fact of their expense (Sunstein 2005)? Some forms of risk-aversion might reflect limited attention or bounded rationality. There is a great deal of thinking to be done about when people are risk-averse and when they are not, and about when and whether regulators should take account of people's risk preferences in imposing, or in failing to impose, regulations.

## **Welfare Now**

The United States Constitution is the oldest surviving constitution in the world—236 years and counting. Circular A-4 cannot claim that degree of longevity, but in light of the rapid movement of relevant political winds and the high stakes, 20 years is a long time for a document that sets out fundamental principles of regulatory analysis. There is ample room for discussion of the new Circular A-4 and economists will be central to the conversation. After all, a constitution establishes general rules, but considerable space remains to argue about how those rules should be interpreted in specific contexts.

A number of topics that deserve additional exploration have been mentioned already. Here, I would emphasize as starting points the work of Gabaix (2019) on behavioral inattention, which may help organize diverse behavioral findings; the work of Adler (2019) on thinking about distributional issues in the regulatory context through the lens of prioritarianism; and the work of Thunström (2019) on the complex effects of disclosure policies, which may affect different people in different ways, and which may even hurt rather than help people with self-control problems. In addition (and this remains a potentially serious gap), no president has yet authorized or directed agencies to focus on subjective welfare, measured by reference to actual experience (as discussed in this journal by Kahneman and Krueger 2006), though the topic has occasionally been discussed at high levels. Finally, the Executive Office of the President itself has drawn attention to an assortment of “frontiers” issues, on which much further work needs to be done (National Science and Technology Council 2023). Those issues include nonfatal health effects, ecosystem services, wildfire and extreme weather effects, and the effects of information provision and transparency requirements. There is a great deal of room for both theoretical and empirical work on those issues.

Existing Regulatory Impact Analyses, both proposed and final, are generally in the public domain, and they might provide the foundation for that work. The Economic Constitution of the United States remains a work in progress.

■ *The author thanks the editors of this journal for exceptionally valuable comments and suggestions, and also thanks David Olin for superb research assistance.*



## References

- Adler, Matthew D.** 2016. "Benefit–Cost Analysis and Distributional Weights: An Overview." *Review of Environmental Economics and Policy* 10 (2): 264–85.
- Adler, Matthew D.** 2019. *Measuring Social Welfare: An Introduction*. Oxford: Oxford University Press.
- Adler, Matthew D., and Ole F. Norheim, eds.** 2022. *Prioritarianism in Practice*. Cambridge, UK: Cambridge University Press.
- Adler, Matthew D., and Eric A. Posner.** 2006. *New Foundations of Cost-Benefit Analysis*. Cambridge, MA: Harvard University Press.
- Allcott, Hunt, and Judd B. Kessler.** 2019. "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons." *American Economic Journal: Applied Economics* 11 (1): 236–76.
- Arrow, Kenneth J., Maureen L. Cropper, Christian Gollier, Ben Groom, Geoffrey Heal, Richard G. Newell, William D. Nordhaus et al.** 2014. "Should Governments Use a Declining Discount Rate in Project Analysis?" *Review of Environmental Economics and Policy* 8 (2): 145–63.
- Biden, Joseph.** 2021. "Modernizing Regulatory Review." January 20. <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/modernizing-regulatory-review/>.
- Breyer, Stephen.** 1980. *Regulation and Its Reform*. Cambridge, MA: Harvard University Press.
- Caplin, Andrew, and John Leahy.** 2004. "The Social Discount Rate." *Journal of Political Economy* 112 (6): 1257–68.
- Carlson, Deven, Simon Haeder, Hank Jenkins-Smith, Joseph Ripberger, Carol Silva, and David Weimer.** 2019. "Monetizing Bowser: A Contingent Valuation of the Statistical Value of Dog Life." *Journal of Benefit-Cost Analysis* 11 (1): 131–49.
- Chetty, Raj.** 2015. "Behavioral Economics and Public Policy: A Pragmatic Perspective." *American Economic Review* 105 (5): 1–33.
- Cowen, Tyler, and Derek Parfit.** 1992. "Against the Social Discount Rate." In *Justice between Age Groups and Generations*, edited by James S. Fishkin and Peter Laslett, 144–61. New Haven, CT: Yale University Press.
- Environmental Protection Agency (EPA).** 2023. "EPA Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances." [https://www.epa.gov/system/files/documents/2023-12/epa\\_scghg\\_2023\\_report\\_final.pdf](https://www.epa.gov/system/files/documents/2023-12/epa_scghg_2023_report_final.pdf).
- Exec. Order No. 12291**, 46 Fed. Reg. 13193 (February 17, 1981).
- Exec. Order No. 12866**, 58 Fed. Reg. 51735 (October 4, 1993).
- Exec. Order No. 13563**, 76 Fed. Reg. 3821 (January 21, 2011).
- Exec. Order No. 13771**, 82 Fed. Reg. 9339 (January 30, 2017).
- Exec. Order No. 14094**, 88 Fed. Reg. 21879 (Apr. 11, 2023).
- Fleurbay, Marc, and Rossi Abi-Rafeh.** 2016. "The Use of Distributional Weights in Benefit–Cost Analysis: Insights from Welfare Economics." *Review of Environmental Economics and Policy* 10 (2): 286–307.
- Frank, Robert H.** 2005. "Positional Externalities Cause Large and Preventable Welfare Losses." *American Economic Review* 95 (2): 137–41.
- Gabaix, Xavier.** 2019. "Behavioral Inattention." In *Handbook of Behavioral Economics: Applications and Foundations*, Vol. 2, edited by B. Douglas Bernheim, Stefano DellaVigna and David Laibson, 261–343. Amsterdam: North-Holland.
- Gilovich, Thomas, Dale Griffin, and Daniel Kahneman.** 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge, UK: Cambridge University Press.
- Harberger, Arnold C.** 1978. "On the Use of Distributional Weights in Social Cost-Benefit Analysis." *Journal of Political Economy* 86 (2): S87–120.
- Heal, Geoffrey.** 2007. "Discounting: A Review of the Basic Economics." *University of Chicago Law Review* 74 (1): 59–77.
- Hemel, Daniel.** 2022. "Regulation and Redistribution with Lives in the Balance." *University of Chicago Law Review* 89 (3): 649–734.
- Kahneman, Daniel, and Alan B. Krueger.** 2006. "Developments in the Measurement of Subjective Well-Being." *Journal of Economic Perspectives* 20 (1): 3–24.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky, eds.** 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.
- Kaplow, Louis, and Steven Shavell.** 2002. *Fairness versus Welfare*. Cambridge, MA: Harvard University Press.

- Kotchen, Matthew J.** 2018. "Which Social Cost of Carbon? A Theoretical Perspective." *Journal of Association of Environmental and Resource Economics* 5 (3): 673–93.
- Lavetti, Kurt.** 2023. "Compensating Wage Differentials in Labor Markets: Empirical Challenges and Applications." *Journal of Economic Perspectives* 37 (3): 189–212.
- Millner, Antony, and Geoffrey Heal.** 2023. "Choosing the Future: Markets, Ethics, and Rapprochement in Social Discounting." *Journal of Economic Literature* 61 (3): 1037–87.
- National Science and Technology Council.** 2023. *Advancing the Frontiers of Benefit-Cost Analysis: Federal Priorities and Directions for Future Research*. Washington, DC: US Government Printing Office.
- Nordhaus, William.** 2007. "Critical Assumptions in the Stern Review on Climate Change." *Science* 317 (5835): 201–02.
- Nurmi, Väinö, and Heini Ahtainen.** 2018. "Distributional Weights in Environmental Valuation and Cost-Benefit Analysis: Theory and Practice." *Ecological Economics* 150: 217–28.
- Office of Management and Budget (OMB).** 2003. Circular A-4. [https://obamawhitehouse.archives.gov/omb/circulars\\_a004\\_a-4/](https://obamawhitehouse.archives.gov/omb/circulars_a004_a-4/).
- Office of Management and Budget (OMB).** 2023a. "Preamble: Proposed Circular No. A-4, 'Regulatory Analysis.'" <https://www.whitehouse.gov/wp-content/uploads/2023/04/DraftCircularA-4Preamble.pdf>.
- Office of Management and Budget (OMB).** 2023b. "Draft for Public Review: Circular A-4." <https://www.whitehouse.gov/wp-content/uploads/2023/04/DraftCircularA-4.pdf>.
- Office of Management and Budget (OMB).** 2023c. Circular A-4. <https://www.whitehouse.gov/wp-content/uploads/2023/11/CircularA-4.pdf>.
- Office of Management and Budget (OMB).** 2023d. "Appendix: Default Social Rate of Time Preference Estimates." <https://www.whitehouse.gov/wp-content/uploads/2023/11/CircularA-4Appendix.pdf>.
- Rabin, Matthew.** 2013. "Healthy Habits: Some Thoughts on the Role of Public Policy in Healthful Eating and Exercise under Limited Rationality." In *Behavioural Public Policy*, edited by Adam Oliver, 115–47. Cambridge: Cambridge University Press.
- Robinson, Lisa A., William J. Raich, James K. Hammitt, and Lucy O'Keefe.** 2019. "Valuing Children's Fatality Risk Reductions." *Journal of Benefit-Cost Analysis* 10 (2): 156–77.
- Ruggari, Kai, Sonia Ali, Mari Louie Berge, Giulia Bertoldo, Ludvig D. Bjørndal, Anna Cortijos-Bernabeu, Clair Davison et al.** 2020. "Replicating Patterns of Prospect Theory for Decision under Risk." *Nature Human Behavior* 4: 622–33.
- Samuelson, William, and Richard Zeckhauser.** 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty* 1 (1): 7–59.
- Sen, Amartya.** 2000. "The Discipline of Cost-Benefit Analysis." *Journal of Legal Studies* 29 (S2): 931–52.
- Sunstein, Cass R.** 2004. "Lives, Life-Years, and Willingness to Pay." *Columbia Law Review* 104 (1): 205–52.
- Sunstein, Cass R.** 2005. *Laws of Fear: Beyond the Precautionary Principle*. Cambridge, UK: Cambridge University Press.
- Sunstein, Cass R.** 2014. "The Limits of Quantification." *California Law Review* 102 (6): 1369–1421.
- Sunstein, Cass R.** 2018. *The Cost-Benefit Revolution*. Cambridge, MA: The MIT Press.
- Sunstein, Cass R.** 2020. *Too Much Information: Understanding What You Don't Want to Know*. Cambridge, MA: The MIT Press.
- Sunstein, Cass R.** 2021a. *Averting Catastrophe*. New York: New York University Press.
- Sunstein, Cass R.** 2021b. *Sludge: What Stops Us from Getting Things Done and What to Do about It*. Cambridge, MA: The MIT Press.
- Thaler, Richard H.** 1994. "Psychology and Savings Policies." *American Economic Review* 84 (2): 186–92.
- Thaler, Richard H.** 2016. "Behavioral Economics: Past, Present, and Future." *American Economic Review* 106 (7): 1577–1600.
- Thunström, Linda.** 2019. "Welfare Effects of Nudges: The Emotional Tax of Calorie Menu Labeling." *Judgment and Decision Making* 14 (1): 11–25.
- Tversky, Amos, and Daniel Kahneman.** 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5 (2): 207–32.
- Viscusi, W. Kip.** 1993. "The Value of Risks to Life and Health." *Journal of Economic Literature* 31 (4): 1912–46.
- Viscusi, W. Kip.** 1995. *Fatal Tradeoffs: Public and Private Responsibilities for Risk*. Oxford: Oxford University Press.

- Viscusi, W. Kip.** 2010. "The Heterogeneity of the Value of Statistical Life: Introduction and Overview." *Journal of Risk and Uncertainty* 40: 1–13.
- Viscusi, W. Kip, and Joseph E. Aldy.** 2003. "The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World." *Journal of Risk and Uncertainty* 27 (1): 5–76.
- Viscusi, W. Kip.** 2022. "Efficiency Criteria for Nudges and Norms." *Public Choice* 191: 465–82.
- Weisbach, David A.** 2015. "Distributionally Weighted Cost–Benefit Analysis: Welfare Economics Meets Organizational Design." *Journal of Legal Analysis* 7 (1): 151–82.
- Weitzman, Martin L.** 1994. "On the 'Environmental' Discount Rate." *Journal of Environmental Economics and Management* 26 (2): 200–209.
- Weitzman, Martin L.** 1998. "Why the Far-Distant Future Should Be Discounted at Its Lowest Possible Rate." *Journal of Environmental Economics and Management* 36 (3): 201–08.
- Williams, Sean.** 2013. "Statistical Children." *Yale Journal on Regulation* 30: 63–124.x



# The Financial Crisis Inquiry Commission and Economic Research

Wendy Edelberg and Greg Feldberg

**T**he financial and economic crises of 2007–2009 were extraordinarily disruptive. They created a “Great Recession” that left millions unemployed and destroyed trillions of dollars in household wealth. The economic recovery was painfully slow, and the financial system still bears the scars. In May 2009 (as it turned out, one month before the trough of the recession), the US Congress created the Financial Crisis Inquiry Commission (FCIC) to conduct an independent investigation of the causes of the crisis. That investigation served as the evidence base for public hearings, internal and external reports, and in January 2011, a 633-page *Final Report* (Financial Crisis Inquiry Commission 2011). The report almost certainly remains the most cited source about the crisis in articles, books, and court cases; it is still used extensively in university classes in economics, policy, business, and law departments; and the results of our investigation helped inform the US Department of Justice and the Securities and Exchange Commission in fraud cases that brought in billions of dollars in fines from financial institutions.

Our investigation was shaped in substantial part by economists and other experts. Some of those experts were on our staff. Staffing was a challenge given

■ *Wendy Edelberg is Director, Hamilton Project, and Senior Fellow, Economic Studies, both at the Brookings Institution, Washington, DC. She is also a Principal, Westexec Advisers, Washington, DC. Greg Feldberg is Director of Research for the Yale Program on Financial Stability and Research Scholar at the Yale School of Management, both in New Haven, Connecticut. Edelberg was Executive Director and Feldberg was Director of Research for the Financial Crisis Inquiry Commission, created by the US Congress, which published its Final Report in January 2011. Their email addresses are [wedelberg@brookings.edu](mailto:wedelberg@brookings.edu) and [greg.feldberg@yale.edu](mailto:greg.feldberg@yale.edu).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.2.43>.

the tight deadline, and experts willing to work long hours were welcome. We also benefited from research published before and during the investigation and had numerous interactions with experts who made themselves available as outside resources. We asked many researchers who had published well-cited, data-driven papers to submit testimony, brief Commissioners, and speak with our staff in recorded interviews. We also needed help from researchers who had access to key data sources and understood how to use them.

In this article, we describe ways that economists and other experts can best be a resource for government efforts like the staff work of the Financial Crisis Inquiry Commission. In particular, early research on a new issue is invaluable, even in the fog of confusion and uncertainty. Although we began our work with the FCIC barely one year after the height of the crisis, there was already a strong core of crisis literature, and many aspects of the crisis were well-understood. Central to the financial crisis were subprime mortgages—mortgages to borrowers with a relatively high risk of default—and the securities into which they were packaged. There was little debate that there had been widespread defaults on subprime mortgages, most of which had been packaged, securitized, and sold by banks and investment banks as “private-label mortgage-backed securities”—so named because they were not backed by government-sponsored enterprises. In turn, the riskiest of these securities had been repackaged and resecuritized into other financial products called “collateralized debt obligations,” which were often funded short-term with asset-backed commercial paper and repurchase agreements. There was also little debate about what had happened to these products. Some of the country’s largest financial institutions—including AIG, Merrill Lynch, and Citigroup—had failed or nearly failed after reporting billions of dollars of losses on collateralized debt obligations. Widespread runs on asset-backed commercial paper and repurchase agreements, in which borrowers found themselves unable to roll over short-term debt due to concerns about the mortgage market and the financial health of financial institutions, marked the two most critical episodes of the financial crisis in the summer of 2007 and the fall of 2008. The government had taken extraordinary steps to support many large companies, leading to a resurrection of the term “too-big-to-fail.”

Yet many questions remained unanswered or controversial, due to lack of data or incomplete analysis. There was much debate about how the losses in subprime mortgages had gotten so bad and how they had become concentrated in those too-big-to-fail companies. Were collateralized debt obligations and asset-backed commercial paper accelerants of the crisis, or simply badly designed financial products? What role had been played by derivatives, financial products that mirror the performance of referenced assets, which were ubiquitous in the production and trading of collateralized debt obligations? How had so much of the financial system become dependent on short-term funding? Why had the managers of the country’s largest banks, investment banks, and insurance companies made decisions that in retrospect proved so self-destructive?

The most contentious question of all had to do with government housing policies that subsidized homeownership. Had they played a significant role in the financial shenanigans of the boom years? More to the point, had the government-sponsored enterprises contributed to the debacle around private-label mortgage-backed securities? This question was the source of much division in the contemporaneous public debate. Both sides of the political aisle used the issue as a wedge to argue for or against official interventions in the housing market and, more broadly, the economy. Our empirical and investigative work ultimately concluded that housing policy had not been a leading cause of the crisis, and nine out of ten Commissioners agreed with that conclusion.

Other factors did emerge as important causes, as we describe below. Rather than think of the crisis as *a perfect storm*, with the implication that it was an unfortunate but unavoidable confluence of unpredictable events, the staff sought to distinguish between risks that were firmly in the public debate in the precrisis years of the mid-2000s and those that were not. On the one hand, the media of 2005 and 2006 were replete with discussions of a housing bubble and potential crash. On the other hand, there was little understanding among experts in business, government, or academia that securitization and derivatives could concentrate risks, or that short-term funding markets could be vulnerable to old-fashioned banking runs. Similarly, there was general awareness that “self-regulation” had created a hands-off relationship between regulators and financial institutions, but a poor understanding of how little appreciation some of those institutions had about their own risk exposure.

In this essay, we begin with some background on the Financial Crisis Inquiry Commission itself, including how its work was organized and staffed. We then describe four ways economists and other experts contributed to our work: joining the staff; conducting timely and relevant research; participating in hearings and interviews; and supporting various novel data projects that the staff created to answer questions for which existing data were inadequate. We close with a case study on the internal controversy over the role of government housing policy in the crisis and the analysis we conducted to address it.

## **Origins of the Financial Crisis Inquiry Commission: Mission and Staffing**

The Financial Crisis Inquiry Commission was created by the Fraud Enforcement and Recovery Act of 2009 (Public Law 111–21), with six Commissioners to be appointed by the Congressional majority and four by the minority. The Commission was given subpoena power and authorized to hold hearings. It was required to submit a report to Congress in December 2010 (although the report was actually submitted the following month).

When the enabling legislation was signed into law in May 2009, the Democratic party had a Congressional majority. Thus, the Commission had six members

appointed by the Democratic leadership in Congress and four by the Republican leadership. The six Democratic appointees were Phil Angelides (chair), Brooksley Born, Byron Georgiou, Bob Graham, Heather H. Murren, and John W. Thompson; the four Republican appointees were Bill Thomas (vice chair), Keith Hennessey, Douglas Holtz-Eakin, and Peter J. Wallison.

Once the chair and vice chair hired an executive director, those three people largely organized the approach to the work and allocated the budget, although all of the Commissioners periodically met and weighed in. When the Commissioners met, they also reviewed and debated interim findings and considered how to report those findings to Congress. The Commission was ultimately given a budget of \$9.8 million, which largely went to maintaining a staff of 87 people that rotated through a nondescript floor of an office building near the White House.

From the early days, it was clear that the Financial Crisis Inquiry Commission would operate in a challenging political environment, internally and externally. Since a majority of Commissioners had been appointed by one political party, there was little urgency for all ten Commissioners to seek common ground. In contrast, earlier Congressional commissions such as the 9/11 Commission (National Commission on Terrorist Attacks Upon the United States 2004) and the commission created in the aftermath of the savings and loan crisis of the late 1980s and the early 1990s (National Commission on Financial Institution Reform, Recovery and Enforcement 1993) had equal party representation.

Moreover, the first twelve months of the Commission's work coincided with an intense battle in Congress over the legislative response to the crisis. The senior staff assistant to the vice chair later wrote a PhD thesis on "organizational politics" that incorporated the example of a Democratic Congress that was reluctant to create an independent commission to study the causes of the crisis, lest the conclusions dilute the momentum for its desired financial reforms (Ganz 2016). On the other hand, internal discussions among Commissioners (later made public) show that for at least one Republican Commissioner, the risk was rather that the *Final Report* could support new regulations (Democratic Staff of the Committee on Oversight and Government Reform 2011). At any rate, in July 2010, while the Commission was in the middle of the allotted time for its task, the Democrat-led Congress passed the Dodd-Frank Wall Street Reform and Consumer Protection Act, the most significant financial regulatory reform in decades, with few Republican votes.

The leaders of the Financial Crisis Inquiry Commission hoped to meet the challenge of partisanship through independent and disinterested analysis. The week of his appointment in July 2009, FCIC Chair Phil Angelides told the press that he believed "the mission is so important that it can and must transcend" partisanship (as quoted in Boles 2009). In a similar spirit, Vice Chair Bill Thomas, the senior Republican on the Commission and a former college instructor, said in his opening remarks at an early hearing: "[W]e are the diagnosticians for the U.S. economy; we also have to be investigators, economists and historians. But, most of all, we have to be teachers. . . . In order to be good teachers, we must first be good students" (Thomas 2010).



The mandate to determine the causes of the financial crisis was unusually broad, compared to similar post-crisis initiatives. For example, when Congress had created the National Commission on Financial Institution Reform, Recovery and Enforcement (1993) almost two decades earlier to investigate the savings and loan crisis, whose work the FCIC staff admired and occasionally sought to emulate, the primary focus was on the broad causes of the crisis from a largely academic perspective. Ferdinand Pecora—who actually conducted the legendary investigation of the financial collapse of 1929 in a series of Senate Committee on Banking and Currency (1934) hearings, rather than through a commission—focused on institutional failures in the financial sector, rather than underlying causes. In contrast, the legislative mandate of the Financial Crisis Inquiry Commission required us to examine both the failures or near-failures of major individual financial institutions and the broad causes of the crisis. The law even gave us a list of 22 possible causes to analyze; for example, the first 10 possibilities it listed were fraud, regulatory failure, savings or fiscal imbalances, monetary policy, accounting practices, tax policy, capital and liquidity requirements, credit rating agencies, lending practices and securitization, and affiliations between banks and nonbanks.

The Commissioners decided early to pick targets carefully, recognizing it would be impossible to cover everything about the financial crisis in less than 18 months between their appointment in July 2009 and the December 2010 deadline for the *Final Report*. Time and money were both major constraints. The Commission did not expect to pay equal attention to all 22 topics mentioned in the statute; there would be no ranking of causes. Also, it could not comprehensively investigate every company that played a role in the crisis. For example, for an analysis of the role of credit rating agencies, the Commission chose to focus its investigation on Moody's rather than S&P or Fitch; for the analysis of government-sponsored enterprises, it focused on Fannie Mae rather than Freddie Mac. In some cases, the Commission made these decisions because other entities had already conducted credible investigations of some institutions, and the Financial Crisis Inquiry Commission did not need to duplicate previous work (Ward 2020). We quickly determined that our comparative advantage was in studying the financial crisis. As a result, to the degree the FCIC's *Final Report* opined on the causes of the economic crisis, the report largely reflected contemporaneous academic thinking.

The chair, vice chair, and executive director, with support from the remaining Commissioners, decided to organize the staff into two groups—research and investigation—reflecting the dual mandate from Congress. Senior hires across the staff were approved by the chair and vice chair while the executive director was largely given authority to hire less senior people without approval.

A research team led by the director of research had the primary mission to investigate the underlying causes of the crisis. Once hired, the first responsibility of the director of research was to develop a plan with the executive director, chair, and vice chair for hiring a team and organizing the work. The research staff included PhD economists who specialized in topics like mortgage finance, macroeconomics, and derivatives; academics and practitioners with deep financial and regulatory

experience, many on detail from regulatory agencies such as the Federal Reserve and the Federal Deposit Insurance Corporation; and a few academics brought in on a temporary basis to support the production of specific preliminary reports or chapters in the *Final Report*. Early hires included experts who had strong prior knowledge of the relevant academic literature.

The research team prepared “preliminary staff reports” exploring potential causes of the crisis across a wide range of topics. We published these preliminary reports prior to hearings on related topics.<sup>1</sup> The existing early-stage economic research and interviews with academics were a critical input. In addition, the research team analyzed confidential information collected by the Financial Crisis Inquiry Commission and participated in interviews with financial market participants. The preliminary staff reports and other work by the team helped to create a baseline understanding for FCIC staff and Commissioners of how the financial system came to be what it was in 2008 and where the fragilities were. In addition, the research team highlighted issues that may have more specifically led to institutional failures.

The investigations staff was initially led by three senior investigators hired by the executive director, chair, and vice chair. Staff had predominantly legal backgrounds. The investigative teams were ultimately combined under one director of investigations, promoted from that group. The teams led the production of “preliminary investigation reports,” which were case-study investigations of specific financial firms and regulatory agencies that played pivotal roles in the crisis. These preliminary reports remained confidential at the time, but are now available through the National Archives. The preliminary investigation reports largely reflected the analysis of confidential information related to particular financial institutions and interviews of financial market participants and experts.

There was continual cross-pollination between the research and investigative teams. Members of both teams attended most interviews, promoting a common understanding of the crisis narrative that emerged across the staff over the course of the months we worked together. Research staff contributed to case studies and investigation staff contributed to research reports.

The staff was largely in place by spring 2010, almost a year after the Commission was initially formed. Roughly speaking, staff focused mostly on hearing preparation during the first half of 2010 and mostly on drafting the *Final Report* during the second half; for the report, research staff took the lead on thematic or historic chapters, while investigation staff took the lead on institution-focused chapters. As the focus changed, the composition of the staff changed modestly. For example,

<sup>1</sup>Stanford University’s Rock Center for Corporate Governance and Stanford Law School together host an archival website for the work of the Financial Crisis Inquiry Commission, which includes not just the *Final Report*, but also background documents, information on hearings, emails, audio recordings, and more. The ten “preliminary staff reports” are available at <https://fcic.law.stanford.edu/resource/reports/> (accessed on March 4, 2024). The Yale Program on Financial Stability also hosts an archival copy of the FCIC website at <https://ypfs.som.yale.edu/financial-crisis-inquiry-commission> (accessed on March 4, 2024).

once the focus was less on investigation and more on drafting, some investigation staff left the Commission and people with writing and editing expertise were hired.

Academic researchers on the staff were essential to our work. They helped to interpret outside research, conduct analysis, and communicate findings. However, the work of a government commission can stretch well beyond the typical skill set of an academic researcher. Staff often need to be generalists, use and communicate a range of research findings (not just the individual's preferred findings), work on a deadline, and be comfortable with less autonomy and with publishing analysis amidst substantial uncertainty. As a result, researchers from outside academia, such as those from government agencies, also played an essential role on the staff.

One lesson from our experience is that a government entity is—or should be—hesitant to hire researchers with an established and immovable point of view. Some unsolicited offers to join the staff or write parts of the report were less than helpful. For example, some prominent academic economists offered to commit significant time to the Commission with the stipulation that the findings in their previous research would be featured. Even in cases where the evidence we were gathering supported those findings, we found such agreements too risky. We needed people who, first, had a fully open mind to the unfolding evidence and, second, were willing to subordinate their own agenda and priorities to the larger project. That said, we were able to usefully interact with those academics on an informal basis and they were valuable external resources.

It should be acknowledged that even for those who have the skills and interest, getting hired can be a messy and complicated process. In our case, the Commission did the bulk of its hiring over the course of just a few months. Hiring plans may be already settled by the time researchers become aware of the staff's work. Then, subsequent hiring happened when we assessed an urgent need and looked for someone who could join the staff within just a few weeks. As a result, in the case of a short-lived project like the Commission, the way that researchers become aware of the staff's work and staff become aware of potential hires was often through word-of-mouth.

## **Timely and Relevant Research**

When major economic and financial events happen, at least some researchers will find themselves with the combination of prior expertise, access to data, time, and enthusiasm to pivot from their existing research projects. Experts can make a valuable contribution to the understanding of an emerging issue by conducting innovative research early on, even while recognizing the inevitable limitations of work done so quickly. Early published research from 2008 and 2009 on the financial crisis was critical to the work of the Financial Crisis Inquiry Commission.

From the very start of the Commission's work, the research team sought to develop an understanding of the arguments over the causes of the crisis. The research team pored over early research about the causes of the crisis from sources

like the Federal Reserve Bank of Kansas City (2008, 2009) annual conferences in Jackson Hole, a collection of white papers from Stern Business School faculty at New York University (Acharya and Richardson 2009), reports published by Baily, Litan, and Johnson (2008) and by the Committee on Capital Markets Regulation (2009), and a collection of articles about the financial crisis in a double-issue of *Critical Review*.<sup>2</sup> We also reviewed two symposiums organized by this journal: the Symposium on Early Stages of the Credit Crunch in the Winter 2009 issue and the Symposium on Financial Plumbing in Winter 2010.

We often found valuable contributions outside these high-profile initiatives. In the immediate wake of the financial crisis, there was a strong global demand for new perspectives on an event that many in the economics profession had gotten wrong. Working papers sometimes had an outsized influence. Moreover, the crisis emanated largely from sectors of the financial system where few academic economists had expertise, such as investment banking, securitization, derivatives, and asset management. The term “shadow banking” had become popular to describe such activities that mirrored traditional banking activities, but were done outside the traditional banking sector and thus without the same traditional banking safeguards. New voices from academic-adjacent fields sometimes emerged with original insights on these activities. For example, Pozsar (2008) produced the first version of his very influential “shadow banking map” while at Moody’s Economy.com, before moving to the Federal Reserve Bank of New York. Another example was the Harvard undergraduate who had acquired for her remarkable thesis, and generously shared with us, a database of hundreds of billions of dollars of collateralized debt obligation securities that had channeled much of the risk in the mortgage market in opaque ways (Barnett-Hart 2009). Her work had been part of the background material for the book (and later movie) *The Big Short* (Lewis 2010).

Some experts have built their careers on work they launched in those years, and some working papers had outsized influence amidst the strong demand for insights. Meanwhile, journals invited scholars to write papers on timelines that were much faster than normal. The most effective work of this kind acknowledges the sources of uncertainty and questions that remained unaddressed. Nonetheless, it is an essential start for pointing government entities in the right direction.

Outside experts can also volunteer to be a resource—helping government staff get up to speed very quickly on the state of knowledge, providing insights about where the staff should focus their efforts, and pointing to data and other evidence. To help government staff understand where the academic literature stands on an emerging issue, literature reviews are invaluable. Even if such a review is well-known among academics, directly sharing it with government staff is useful. For a few examples, the Financial Crisis Inquiry Commission relied on reviews of historical developments in financial regulation, financial literacy, and the effect of financial compensation on incentives of managers. We benefited from earlier

<sup>2</sup>The table of contents for this issue of *Critical Review* is available at <https://www.tandfonline.com/toc/rcr20/21/2-3> (accessed on March 4, 2024).

work like Calomiris (2000) and Wilmarth (2002), and followed up with the authors of both. Moreover, the creation of such reviews is a great service when an urgent issue emerges that a government entity is grappling with—say, a financial crisis, a pandemic, or proposed immigration reform. Comprehensiveness, clarity, and speed are all essential for a successful review in such circumstances.

As mentioned above, sometimes unsolicited offers to help were more problematic, including from academic institutions and foundations that wanted to contribute research and time. When the offers were open-ended and came with no strings, the Commission staff often found a way to benefit. For example, one nonprofit research group shared findings related to their own inquiries into aspects of the crisis and acted as a useful sounding board for the staffs' findings. However, other outside potential collaborators had their own goals, their own timing needs, and were committed to particular points of view. As a result, the Commission ultimately turned down several offers to partner with outside institutions.

## **Participation in Hearings and Interviews**

When researchers make public their early-stage research in response to events, it also becomes a very effective way for government staff to identify researchers working on the subject at hand. The Financial Crisis Inquiry Commission conducted a total of eleven public events—both hearings and forums—between October 2009 and September 2010, over a total of 19 days.<sup>3</sup> The first, second, and fourth hearings provided leading academics an opportunity to describe their prior research on the causes of the crisis. In the third hearing, the Commission sought the views of leading financial institution representatives and financial market participants, among others.

In the three events featuring those who had conducted related research, the Commission invited 17 economists to participate, including “roundtable discussions” in October and November 2009 and a two-day “Forum to Explore the Causes of the Financial Crisis” in February 2010. The roundtable discussions were broad-ranging; the two-day forum consisted of nine sessions led by economists on specific aspects of the crisis, most of whom were authors of early research on the causes of the financial crisis. Participants took a variety of approaches to the problem and represented a diversity of political perspectives. We encouraged participants not only to explain their own research findings, but to also describe points of disagreement and where there was a lack of consensus. Taken together, the papers and reports that these researchers had published served as a crash course on the existing crisis literature.

This crash course was valuable not just for the staff; it also served to educate the Commissioners, who came from very different backgrounds. Moreover, insight

<sup>3</sup>For information on the hearings and participants, see the archival website of the Financial Crisis Inquiry Commission website at <https://fcic.law.stanford.edu/hearings> (accessed on March 4, 2024).

on what could be a potential cause was just as important as insight on what avenues would be less fruitful. For example, the economic research suggested greater focus on high levels of leverage and financial contagion across the financial sector as key dynamics of the crisis, and less focus on other potential causes like systemic fraud among homeowners and risks due to models of executive compensation.

The work of the Commission then shifted to focus on the five topical hearings, which represented the core of the Commission's investigation, broadly covering crucial topics: (1) subprime mortgage lending, securitization, and the government-sponsored enterprises; (2) the so-called "shadow banking" system; (3) the credibility of credit ratings agencies; (4) the role of financial derivatives; and (5) the risks posed by "too big to fail" financial institutions. Each of these topical hearings was preceded by six to eight weeks of literature reviews and expert interviews, often with outside researchers, in which both research and investigation staff participated. Staff reached out directly to dozens of researchers who had published papers on the crisis—explaining why investors ran certain markets, why incentive problems pervaded securitization markets, and why risk management failed at so many large companies. These interactions included invitations to submit testimony, to brief Commissioners, and to speak with our staff in recorded interviews. These five topical hearings provided the Commission with many opportunities to put executives from one or more case-study companies and their regulators on the spot, sometimes with panels of economists and other experts to provide context and analysis.

Based on these experiences, we believe that the best way for a researcher to help government staff get up to speed is to be willing to explain not just research that was personally conducted, but more generally explain where the literature is broadly. Government staff need a framework to understand disagreements in the existing literature. A government report is not likely to incorporate the full worldview of a single expert, no matter how compelling. Staff is best-served when researchers can be honest brokers not just for their own research, but also in characterizing where there is and is not consensus and elucidating points of disagreement.

Moreover, for researchers to help staff pursue fruitful avenues for analysis, researchers should be willing to *think outside their models*. Of course, the real world is far more complicated than theoretical models and even empirical models. Recognizing that, staff is best served when researchers can brainstorm about the relevance of their analysis to the problem at hand. Sometimes, government staff need a number—even while appreciating the enormous uncertainty around that number.

## **Novel Data Projects**

Novel data is often the lifeblood of the work done by the staff of a government commission. We were always very aware of the unique historical opportunity that the Financial Crisis Inquiry Commission presented to gather information on the financial crisis, due to its subpoena power and to the fact that the events we were investigating were still fairly recent. In particular, we sought to identify and create

unique datasets that would likely be impossible to create in the future (Ward and Wiggins 2020). The Commission also published the results and the underlying data, appropriately aggregated or anonymized, on its permanent website.

We focused on gathering data about important participants in areas of the financial industry that were largely unregulated and about which financial regulators had limited information prior to the crisis. The staff launched several data projects to support our research.<sup>4</sup> Each project helped us illustrate important elements of the crisis and its causes and contributed significantly to the *Final Report*.

The Commission tried to cast a broad net when looking for available data. For example, as noted, we relied early on for data on collateralized debt obligations on the Harvard undergrad's database until we were able to get a broader confidential dataset from Moody's. A collateralized debt obligation is a financial instrument that bundles other securities and then, much like mortgage-backed securities, sells securities in "tranches" with different credit ratings: the higher-risk tranches would bear any early losses, middle-risk tranches would only bear losses after the higher-risk tranches had lost their value, and the lower-risk tranches would only bear losses when all of the other tranches had lost value. In this way, even if the underlying assets were risky mortgage-backed securities, they could be the basis for a range of tranches ranging from lower to higher risk. Before the crisis, rating agencies gave the lowest-risk tranches their highest (that is, safest) ratings, which turned out to be a terrible mistake when the crisis hit.

Moreover, tranches of collateralized debt obligations based on pools of mortgage-backed securities were then combined into new collateralized debt obligations, sometimes called "CDO-squareds," which in turn could then be divided into new tranches of differing risks. Eventually, the Commission gathered confidential data from Moody's that we used to create charts that illustrated the explosion of collateralized debt obligations and "CDO-squareds" in the lead-up to the financial crisis in 2006 and 2007. We also created on the Commission website a CDO Library, which included pitch books, term sheets, and offering circulars for selected collateralized debt obligations, some of which were subjects of our investigations.

Three of the Commission's major data initiatives were a case study on a mortgage-backed security, a survey of hedge funds, and a survey of short-term funding market participants.

### **Case Study on a Representative Private-Label Mortgage-Backed Security**

Long before we began our work, many sources had described aspects of the "originate-to-distribute" model in which lenders extended credit with the goal of reselling the mortgages to others who would package them into securities for investors. Because the lender would no longer bear any risk of default, this model made it more likely that lenders would extend mortgages to people who would be unable to pay. A paper by two New York Fed economists described "seven deadly frictions" in

<sup>4</sup>For an overview of these projects, see the archival website of the Financial Crisis Inquiry Commission at <https://fcic.law.stanford.edu/resource/staff-data-projects> (accessed on March 4, 2024).

the subprime mortgage market—for example, the incentives for predatory lending and the possibility of adverse selection by asset managers picking high-interest but high-risk mortgages for the securities they sold (Ashcraft and Schuermann 2009).

In an effort to illustrate these issues, staff meticulously tracked an individual mortgage-backed security from the origination of the underlying mortgages to its sale in tranches to buyers, including collateralized debt obligations and the buyers of their tranches. We collected and posted on our website dozens of documents about the deal—for example, the commitment letter from the securitizing bank, due diligence reports, prospectuses and term sheets for marketed securities, internal emails, wire instructions, and confidential emails between the bankers and rating agencies and other parties.<sup>5</sup> The project allowed us to illustrate in the *Final Report* on a very micro level some of the major themes we were exploring with market participants and researchers on a macro level; in particular, how complex, opaque, and open to abuse the process of securitizing and resecuritizing mortgage assets had become in the early 2000s (Cardona and Wiggins 2020). The case study on a representative subprime mortgage-backed securities could not have been completed without help from a former Fed economist.

### **Survey of Hedge Funds**

We collected detailed information from 170 hedge funds with \$1.2 trillion of managed assets, more than half the hedge fund universe by assets at the time. We found strong evidence of two phenomena that market participants and researchers had highlighted in interviews, but for which there was very little existing data.<sup>6</sup> First, we found that 20 percent of hedge fund investors, based on assets, requested redemptions from their funds in the fourth quarter of 2008, contradicting assumptions that hedge funds have access to “sticky money” that allows them to ride out a crisis. We also found evidence of runs by hedge funds on their own bankers, that is, runs on the so-called prime brokers, typically banks or investment banks, with whom they leave their cash.

Second, we found widespread “correlation trading,” defined for our purposes as the purchase of one tranche of a collateralized debt obligation security while shorting other tranches of the same or similar securities with “credit default swaps.” Credit default swaps are a type of derivative product, which market participants can use to take a position on the likelihood that the underlying loans will default on their payments.

Correlation trading was an investment strategy with the goal of making outsized returns if large numbers of subprime borrowers defaulted on their mortgages. The logic was straightforward: if the equity or junior (higher-risk) tranches lost value because of widespread mortgage defaults, it would also increase the risk of the

<sup>5</sup>For details, see the archival website of the Commission at <https://fcic.law.stanford.edu/resource/staff-data-projects/story-of-a-security> (accessed on March 4, 2024).

<sup>6</sup>For details, see the archival website of the Commission at <https://fcic.law.stanford.edu/resource/staff-data-projects/hedge-fund-survey> (accessed on March 4, 2024).



senior and lower-risk tranche. The trader would sell senior tranches short, betting that systemic losses in the mortgage sector would be so bad that they would reach the senior tranches; they would also buy equity tranches to earn a high income in the meantime, while they waited for systemic losses to become apparent.

Market participants had described the practice of correlation trading in interviews with Commission staff. We conducted the survey to find out if the trade was common enough to affect the market. We found that it was. Based on the results of our survey, we estimated that more than half of the equity tranches of all the collateralized debt obligations issued in the second half of 2016 were purchased by funds that also shorted other collateralized debt obligation tranches through credit default swaps. Undoubtedly, this strategy turned out to be very profitable for many investors. However, we concluded that these types of trades also had a deleterious effect on incentives in the structured finance market: “Investors in the equity and most junior tranches of collateralized debt obligations and mortgage-backed securities traditionally had the greatest incentive to monitor the credit risk of an underlying portfolio. With the advent of credit default swaps, it was no longer clear who—if anyone—had that incentive” (Financial Crisis Inquiry Commission 2011, p. 192).

Acquiring data from hedge fund managers about their positions and the activities of their clients during the crisis was not easy, even with our subpoena power. To reassure managers that the information they provided would be strictly confidential, we contracted with the National Opinion Research Center (NORC) at the University of Chicago to collect responses, clean and compile data, and present only aggregate data to FCIC staff. We had no access to individual responses.

### **Survey of Market Participants about Short-Term Funding during Runs on Major Financial Institutions**

We collected data from market participants in short-term funding markets—particularly the markets for commercial paper and repurchase agreements, or repos. Commercial paper was traditionally unsecured and issued by corporations for immediate financial needs. However, during the years before the financial crisis, financial institutions had begun to use commercial paper extensively to provide long-term financing for mortgage-backed securities and collateralized debt obligations. In a repurchase agreement, a borrower sells a security for a short-term, often overnight, but agrees to repurchase it at a slightly higher price the next day. Many financial institutions had been relying on repos to finance large portions of their balance sheets, assuming they would always be able to roll over the previous debt into new debt.

During the financial crisis, as investors and counterparties suddenly became skittish about risks in mortgage-backed securities and financial institutions themselves, it suddenly became costly or impossible to roll over these short-term securities. To those financial institutions, that withdrawal of financial lending felt very much like a traditional bank run. However, the risk that it would become impossible to roll over short-term lending had not been properly understood by market participants, regulators, or researchers before the financial crisis.

We wanted to understand when and how these runs took place, and by whom. We received responses from 17 primary dealers, 18 insurance companies, 19 money market funds, and 30 other mutual funds. In this case, we did not need to call on the National Opinion Research Center to anonymize participants, as in the hedge fund survey. Instead, FCIC staff cleaned and compiled data from individual responses, calculated summary statistics for each of the four types of financial institutions, and published the results on the website.<sup>7</sup> We were able to show the extent of runs on short-term funding markets in summer 2007, March 2008, and September 2008. For example, on average, funds withdrew more than three-quarters of their outstanding repo loans to Lehman Brothers in the week leading up to its failure.

Availability of data about these so-called “shadow-banking” markets has increased since the Financial Crisis Inquiry Commission was active. However, it is likely that a future crisis will one day again emanate from an area that market participants and regulators are not carefully monitoring. A government commission can serve as a lasting resource in identifying and using novel datasets.

## **Politics and the Commission: The Dispute over Government Housing Policy**

No government commission can be divorced from politics, and the Financial Crisis Inquiry Commission was no exception. On the other hand, a Commission that cannot take a stand and emphasize some explanations while rejecting others is not much use. For our Commission, by far the most divisive question was whether government housing policy had been a leading cause of the financial crisis. Typically, the term “housing policy” refers to policies that seek to promote affordable home ownership. It includes, most prominently, the government-sponsored enterprises involved in expanding mortgage finance by securitizing loans and the 1977 Community Reinvestment Act, which encourages banks to make mortgage loans in lower- and middle-income neighborhoods in their communities. Although the role of housing policy was not one of the 22 topics that our enabling legislation specifically identified for our work—the list included, quite broadly, items on “Federal and State financial regulators,” “the legal and regulatory structure of the United States housing market,” and the role of “financial institutions and government-sponsored enterprises”—these issues were already hotly debated in the public arena and had taken on extraordinary significance for those inclined to argue against government interventions of any kind in market outcomes.

The Commission staff grappled with research around the role of housing policy as a cause of the financial crisis. This case study shows the pitfalls of taking all relevant research at face value. Moreover, when different research findings directly contradict each other, future Commissions may have to determine which set of

<sup>7</sup>For discussion, see the FCIC archival website at <https://fcic.law.stanford.edu/resource/staff-data-projects/market-risk-survey> (accessed on March 4, 2024).

findings is more credible. In addition, staff may have to come to such determinations in the face of political disagreements and determined lobbying by researchers.

Early on, the question of housing policy came up in the Commission's discussions with researchers and market participants; in particular, one Commissioner at the outset contended housing policy was the singular cause of the crisis. That Commissioner relied heavily on the work of one researcher in particular, a former Fannie Mae executive named Edward Pinto. In March 2010, Pinto sent a 54-page memo to the Commission describing his theory that federal housing policy was the singular cause of the financial and economic crisis. Pinto (2010) argued that 27 million mortgages, or nearly one-half of the total, were subprime or otherwise low-quality prior to the crisis, and that a significant amount of these low-quality mortgages were driven by these housing policies. In short, his argument was that "most [subprime or other weak loans] were either held by or guaranteed [by] the GSEs or by agencies of the US government."

Commission staff reviewed existing economic research to test Pinto's claim. In a preliminary staff report published for the Commission's hearing on subprime lending on April 7, 2010, FCIC Staff (2010) cited a study by two Federal Reserve economists on the impact of Community Reinvestment Act affordable lending requirements. The study used data that the Fed collects annually under the Home Mortgage Disclosure Act. The data showed that only 6 percent of higher-priced (and therefore likely more risky) loans appeared to be covered by the Community Reinvestment Act. In other words, according to granular data about the housing market, most relatively risky loans were not extended by banks to meet requirements of that law.

However, not all of the Commissioners were satisfied, and they continued to debate Pinto's research internally. Commission staff found we needed to conduct our own analysis to investigate Pinto's broader claim about the extent of low-quality loans in securities backed by the government-sponsored enterprises. As noted, there was little debate that there had been widespread defaults on private-label mortgage-backed securities. The politicized question was whether housing policy had played a large role in the failure of private-label mortgage-backed securities. Government-sponsored enterprises were also involved in mortgage securitization, although mortgage securitizations backed by the government-sponsored enterprises were very different from private-label mortgage-backed securities created by banks and investment banks. The government-sponsored enterprises had a much higher standard for the mortgages they allowed banks to include in these securities. But how much higher?

To examine that question, we sought data from Federal Reserve economists who had access to a sample of loan-level mortgages from the two leading proprietary vendors at the time (Borzekowski and Edelberg 2010a). The sample was very large, representing about 60 percent of the 55 million loans outstanding in the United States at the time. Using aggregate statistics provided by the Federal Reserve economists, Commission staff compared the actual performance of mortgage loans that had been securitized by the government-sponsored enterprises with loans in

private-label mortgage-backed securities. The data allowed us to control for factors that Pinto used to identify low-quality loans, such as low FICO credit ratings or high loan-to-value ratios (in other words, loans that were large relative to the value of the underlying properties) (Borzekowski and Edelberg 2010a, b). Both the staff analysis and Pinto's analysis included Federal Housing Authority (FHA) loans in the category of government-sponsored enterprises.

We found that loans securitized by the government-sponsored enterprises performed substantially better than similar loans securitized in private-label mortgage-backed securities. In our sample, the four million loans backed by government-sponsored enterprises that had FICO credit scores below 660 had an average serious delinquency rate of 6.2 percent in 2008; in comparison, the five million loans in private-label securitizations with FICO scores below 660 had an average serious delinquency rate of over 28.3 percent. Similarly, loans backed by government-sponsored enterprises with loan-to-value ratios above 90 percent had an average serious delinquency rate of 5.7 percent, compared to 15.5 percent for similar loans in private-label securitizations.

Overall, we found that loans placed in private-label mortgage-backed securitizations accounted for the vast majority of serious delinquencies in the mortgage market during the financial crisis; conversely, loans securitized through the government-sponsored enterprises performed perhaps surprisingly well. The staff concluded: “[W]ith the benefit of the performance data from the Federal Reserve (proprietary data that Pinto did not have access to when he completed his analysis), we see that Pinto’s analysis takes millions of loans from the GSE (and FHA) categories that . . . have performed relatively well—and groups them with the subprime and Alt-A securitized loans . . . that have performed relatively poorly” (Borzekowski and Edelberg 2010b).

The Commission’s *Final Report* included a majority report signed by the six Democratic Commissioners, one minority dissent signed by three of the four Republicans, and a second minority dissent signed by a single Republican Commissioner. The majority referred specifically to the conclusions of the staff’s analysis to support its view that housing policy was not a major cause of the financial crisis. The three-person minority dissent emphasized that a wide variety of countries experienced credit booms, housing bubbles, and financial crisis, from which it similarly concluded that “U.S. housing policy is by itself an insufficient explanation of the crisis.” The remaining dissent was written by the Commissioner who had highlighted Pinto’s research throughout the Commission’s work. That Commissioner wrote that housing policy was the central cause of the crisis, repeating Pinto’s view that “*the sine qua non* of the financial crisis was U.S. government housing policy, which led to the creation of 27 million subprime and other risky loans.” His dissent did not refer to the staff’s analysis.

At the last minute, politics threatened to interfere with the release of the Commission’s *Final Report*. Back in summer 2010, as Congress debated additional funding for the Commission on top of the initial \$8 million budget, the senior Republican on the House Oversight Committee announced an investigation into

its activities (Issa 2010). In January 2011, when control of Congress shifted from Democrats to Republicans, that investigation stepped into a higher gear, issuing subpoenas and interview requests to Commissioners and key staff. Ultimately, the new leadership of the House Oversight Committee held no hearings and issued no reports, and the Commission report was released. However, the House Oversight Committee's minority staff did issue a report (Democratic Staff of the Committee on Oversight and Government Reform 2011). Their report describes the internal controversy among Commissioners over Pinto's memos and the staff's response. Emails show the lone dissenting Commissioner urging fellow Republican Commissioners not to say anything that might undermine the effort of House Republicans to modify or repeal the Dodd-Frank Act.

In the end, Commissions established by a political process are inherently subject to political pressures. It is the job of senior leadership within the Commission staff to protect the integrity of the work. No piece of research should be taken at face value. No researchers, regardless of their prominence, should be deferred to. Taking on this responsibility is not for the faint of heart.

### **The Value of Government Commissions: Immediate and Time Capsule**

Many research projects take years to complete. In contrast, the Financial Crisis Inquiry Commission had about 18 months from the passage of the enabling legislation to the publication of the *Final Report*. Moreover, it was working in the immediate aftermath of the Great Recession, in a situation that combined recent and traumatic events, imperfect and limited data, and strong political cross-currents. Nonetheless, the report the Commission produced was valuable, and future Commissions will no doubt produce similarly useful reports. Such reports serve both immediate and long-term purposes. The immediate purpose is to bring to the forefront a broad array of expertise and analysis, to provide a basis for understanding and policy changes. But in addition, a government commission report also serves as a time capsule for what was known at the time; it defines issues and collects data in a way that helps to point later researchers in fruitful directions.

■ *We would like to thank Ron Borzekowski, Erik Hurst, Nina Pavcnik, Timothy Taylor, and Heidi Williams for helpful comments on an earlier draft.*

## References

- Acharya, Viral V., and Matthew Richardson, eds.** 2009. *Restoring Financial Stability: How to Repair a Failed System*. Hoboken, NJ: John Wiley and Sons.
- Ashcraft, Adam B., and Til Schuermann.** 2009. "The Seven Deadly Frictions of Subprime Mortgage Credit Securitization." In *The First Credit Market Turmoil of the 21st Century Implications for Public Policy*, edited by Douglas D. Evanoff, Philipp Hartmann, and George G. Kaufman, 325–43. Singapore: World Scientific.
- Baily, Martin Neil, Robert E. Litan, and Matthew S. Johnson.** 2008. "The Origins of the Financial Crisis." Brookings Institution Fixing Finance Series Paper 3.
- Barnett-Hart, Anna Katherine.** 2009. "The Story of the CDO Market Meltdown: An Empirical Analysis." Harvard Kennedy School, Mossavar-Rahmani Center for Business and Government Associate Working Paper 4.
- Boles, Corey.** 2009. "Financial Crisis Commission Chair: Will Leave No Stone Unturned." *Dow Jones Newswires*, July 15. [https://www.hks.harvard.edu/sites/default/files/centers/mrcbg/files/Barnett-Hart\\_2009.pdf](https://www.hks.harvard.edu/sites/default/files/centers/mrcbg/files/Barnett-Hart_2009.pdf).
- Borzekowski, Ron, and Wendy Edelberg.** 2010a. "Analysis of Housing Data." Memorandum to Commissioners, July 7. <https://ypfsresourcelibrary.blob.core.windows.net/fcic/fcic-docs/2010-07-07%20FCIC%20Staff%20Analysis%20of%20Housing%20Data.pdf> (accessed March 4, 2024).
- Borzekowski, Ron, and Wendy Edelberg.** 2010b. "Analysis of Housing Data and Comparison with Ed Pinto's Analysis." Memorandum to Commissioners, August 21. <https://ypfsresourcelibrary.blob.core.windows.net/fcic/fcic-docs/2010-08-09%20FCIC%20Staff%20Analysis%20of%20Housing%20Data%20and%20Comparison%20with%20Ed%20Pinto%20Analysis.pdf> (accessed March 4, 2024).
- Calomiris, Charles W.** 2000. *U.S. Bank Deregulation in Historical Perspective*. Cambridge, UK: Cambridge University Press.
- Cardona, Mercedes, and Rosalind Z. Wiggins.** 2020. "Lessons Learned: Ron Borzekowski." *Journal of Financial Crises* 2 (4): 55–57.
- Committee on Capital Markets Regulation.** 2009. *The Global Financial Crisis: A Plan for Regulatory Reform*. Cambridge, MA: Committee of Capital Markets Regulation.
- Democratic Staff of the Committee on Oversight and Government Reform** 2011. *An Examination of Attacks against the Financial Crisis Inquiry Commission*. Washington, DC: US House of Representatives.
- FCIC Staff.** 2010. *The Community Reinvestment Act and the Mortgage Crisis*. Preliminary Staff Report, April 7. [https://fcic-static.law.stanford.edu/cdn\\_media/fcic-reports/2010-0407-Preliminary\\_Staff\\_Report\\_-\\_CRA\\_and\\_the\\_Mortgage\\_Crisis.pdf](https://fcic-static.law.stanford.edu/cdn_media/fcic-reports/2010-0407-Preliminary_Staff_Report_-_CRA_and_the_Mortgage_Crisis.pdf) (accessed March 4, 2024).
- Federal Reserve Bank of Kansas City.** 2008. *Maintaining Stability in a Changing Financial System: A 2008 Economic Symposium hosted by the Federal Reserve Bank of Kansas City*. <https://www.kansascityfed.org/research/jackson-hole-economic-symposium/maintaining-stability-changing-financial-system> (accessed March 4, 2024).
- Federal Reserve Bank of Kansas City.** 2009. *Financial Stability and Macroeconomic Policy: Symposium hosted by the Federal Reserve Bank of Kansas City*. <https://www.kansascityfed.org/research/jackson-hole-economic-symposium/financial-stability-and-macroeconomic-policy> (accessed March 4, 2024).
- Financial Crisis Inquiry Commission.** 2011. *The Financial Crisis Inquiry Report: Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States*. <https://www.govinfo.gov/app/details/GPO-FCIC> (accessed March 4, 2024).
- Ganz, Scott Cohn.** 2016. "Three Essays on Organizational Politics, Search, and Learning." PhD diss. Stanford University.
- Issa, Darrell.** 2010. "As Congress Prepares to Pass Additional Funding for FCIC, Issa Questions Financial Commission's Spending and Management," US House Committee on Oversight and Accountability, July 27. <https://oversight.house.gov/as-congress-prepares-to-pass-additional-funding-for-fcic-issa-questions-financial-commissions-spending-and-management> (accessed March 4, 2024).
- Lewis, Michael.** 2010. *The Big Short: Inside the Doomsday Machine*. New York: W. W. Norton.
- National Commission on Financial Institution Reform, Recovery and Enforcement.** 1993. *Origins and Causes of the S&L Debacle: A Blueprint for Reform*. Washington, DC: National Commission on Financial Institution Reform, Recovery and Enforcement.

- National Commission on Terrorist Attacks upon the United States.** 2004. *The 9/11 Commission Report*. <https://govinfo.library.unt.edu/911/report/index.htm> (accessed March 4, 2024).
- Pinto, Edward.** 2010. "Triggers of the Financial Crisis." Memorandum to Staff of FCIC, March 15. [https://yfpsresource.library.blob.core.windows.net/fcic/fcic-docs/2010-03-15%20Triggers%20of%20the%20Financial%20Crisis%20\(Pinto%20memo%20revised\).pdf](https://yfpsresource.library.blob.core.windows.net/fcic/fcic-docs/2010-03-15%20Triggers%20of%20the%20Financial%20Crisis%20(Pinto%20memo%20revised).pdf) (accessed March 4, 2024).
- Pozsar, Zoltan.** 2008. "The Rise and Fall of the Shadow Banking System." *Regional Financial Review*, Moody's Economy.com, July, 13–25.
- Senate Committee on Banking and Currency.** 1934. *Stock Exchange Practices*. Washington, DC: US Government Printing Office.
- Thomas, Bill.** 2010. "Opening Remarks at the Public Hearing of the Financial Crisis Inquiry Commission." Speech, Public Hearing of the FCIC, Washington, DC, January 13, 2020. [https://fcic-static.law.stanford.edu/cdn\\_media/fcic-testimony/2010-0113-Thomas.pdf](https://fcic-static.law.stanford.edu/cdn_media/fcic-testimony/2010-0113-Thomas.pdf) (accessed March 4, 2024).
- Ward, Sandra.** 2020. "Lessons Learned: Phil Angelides." *Journal of Financial Crises* 2 (4): 36–39.
- Ward, Sandra, and Rosalind Z. Wiggins.** 2020. "Lessons Learned: Greg Feldberg." *Journal of Financial Crises* 2 (4): 50–54.
- Wilmarth, Arthur E., Jr.** 2002. "The Transformation of the U.S. Financial Services Industry, 1975–2000: Competition, Consolidation, and Increased Risks." *University of Illinois Law Review* 2002 (2): 215–476.





# Philanthropic Cause Prioritization

Emily Oehlsen

A few years ago, a systematic data collection effort found that philanthropic grantmakers from 157,064 foundations in 23 countries were disbursing over \$150 billion in funding annually (Johnson 2018). Many foundations decide how much and where to give based on their founders' personal precommitments to specific issues, geographies, and/or institutions. If a grantmaking organization instead wanted to select problems based on a general measure of impact per dollar spent, how should it approach this goal? What tools could it use to identify promising cause areas (climate change, education, or health, for example) or to compare grants that achieve different results? These are thorny—and ultimately, largely philosophical—questions with many plausible answers.

This paper focuses on an approach followed by the grantmaking organization Open Philanthropy, my employer, for its “Global Health and Wellbeing” portfolio (as distinct from its work on farm animal welfare and global catastrophic risks).<sup>1</sup> Open Philanthropy has a broad mission: to help others as much as possible with the resources available to it, without any precommitments to particular issues or geographic areas. As a result, we have invested in “cause prioritization,” by which we mean the investigation and selection of focus areas based on their expected net benefits. Open Philanthropy, like many grantmaking organizations, contains “program areas” focused on thematic causes. Once we start giving in a specific cause area, we

■ *Emily Oehlsen is Managing Director, Open Philanthropy, San Francisco, California.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.2.63>.

<sup>1</sup>I will refer to this subset of Open Philanthropy as “Global Health and Wellbeing”, but in practice that name also incorporates work on farm animal welfare. You can read more about Open Philanthropy’s Global Health and Wellbeing portfolio here: <https://www.openphilanthropy.org/our-global-health-and-wellbeing-and-global-catastrophic-risks-grantmaking-portfolios/>.

hire an expert program officer to oversee strategy and grantmaking. From there, we attempt to combine subject-matter expertise with cost-effectiveness analysis. Overall, we seek to benefit from the discipline and insight that quantification and modeling can provide, while taking care to avoid mistaking the legibility of our analysis for ultimate truth. When evaluating a grantmaking opportunity, we use a mix of cost-effectiveness modeling, prompts to elicit strong cases for and against decisions, quantitative predictions, and expert discretion.

Our continually evolving approach consists primarily of two overlapping frameworks. First, we seek to equalize marginal philanthropic returns. Thus, we aim (where possible) to translate the “benefits” of potential grantmaking across different thematic areas addressing premature death, morbidity, and material deprivation into a single unit, then select the highest-impact-per-dollar areas and fund them to a point of equalized marginal benefits. If we believe that an additional dollar can generate more benefits in one cause area versus another, we reallocate. By iteratively comparing incremental opportunities, we aim to arrive at cause areas with marginal opportunities that have similar expected benefits, producing an optimized overall portfolio.

Second, we try to find outlier opportunities by assessing the importance, neglectedness, and tractability of cause areas—what we call “the INT framework.” As an illustrative example, I discuss reducing exposure to lead: it is important enough to cause significant and widespread health and economic harms, which are concentrated among people living in low- and middle-income countries; it is relatively neglected by public and philanthropic funders; and initial evidence suggests tractable interventions in high-burden areas. Although we have not completed our definitive benefit-cost analysis of lead exposure prevention, Open Philanthropy has contributed to early grantmaking focused in this area through our partner GiveWell.<sup>2</sup> Fundamentally, we see the INT criteria as additional proxies for estimating expected cost-effectiveness in areas where doing so directly is too difficult or error-prone, not as a separate set of desiderata.

This essay describes, at a high level, how Open Philanthropy uses the two frameworks of equalizing marginal philanthropic returns and assessing importance, neglectedness, and tractability in practice. The conclusion offers a few reflections on ways we could be wrong.

Our perspective draws upon and applies lessons from many economics-focused philanthropic organizations. Our closest collaborator, GiveWell, uses a very similar

<sup>2</sup>Open Philanthropy began as a partnership between GiveWell and Good Ventures, a philanthropic foundation founded by Cari Tuna and Dustin Moskovitz (who was one of the cofounders of Facebook and Asana). Open Philanthropy became an independent organization in June 2017. For an overview of our history, see <https://www.openphilanthropy.org/about-us/>. Open Philanthropy allocates significant funding based on GiveWell’s recommendations. In July 2021, Open Philanthropy made a grant of \$8 million to Pure Earth (<https://www.givewell.org/research/incubation-grants/Pure-Earth-lead-exposure-july-2021>) and, in December 2021, a grant of \$1.2 million to the Center for Global Development (<https://www.givewell.org/research/incubation-grants/Center-for-Global-Development-lead-exposure-december-2021>), both following GiveWell’s recommendations.

approach. Our work also overlaps with GiveDirectly and the cost-effectiveness standard that direct cash transfers provide, as well as the scaling frameworks used by Development Innovation Ventures at USAID,<sup>3</sup> Evidence Action’s Accelerator,<sup>4</sup> and the Clinton Health Access Initiative (CHAI)’s Incubator.<sup>5</sup> We are avid consumers of economics research (for example, Kremer et al. 2023), we often attempt to synthesize and replicate literatures (for example, Roodman 2023), and we are eager to see further academic economics research that addresses our decision-relevant questions.<sup>6</sup>

While our work at Open Philanthropy is heavily informed by economics, we depart from traditional economics assumptions in a few important ways. As one example, we currently value averting life-years lost equally all over the world, rather than relying on a local willingness to pay. Second, we explicitly depart from the common economics assumption of noncomparability of preferences—that is, we are willing to add up winners and losers and we apply a common (logarithmic) utility function over income for all people in the world. Finally, we generally use normative rather than market or government interest rates for valuing streams of utility.

## Equalizing Marginal Philanthropic Returns

At Open Philanthropy, we seek to apply the basic economic principle of marginal analysis to philanthropic giving. Our Global Health and Wellbeing work is grounded in four claims. First, the world has widespread and cheaply preventable suffering, and a significant portion of it takes the form of premature death, morbidity, and material deprivation among the global poor. Second, the costs and outcomes of interventions tackling these issues are sufficiently similar that they can be usefully compared against one another. Third, when faced with multiple philanthropic opportunities that generate similar benefits, all else equal, we would like to choose the one that produces more benefits per dollar. Finally, most causes exhibit diminishing marginal returns with respect to funding, especially funding from a single source.

Consider the hypothetical example of a grantmaker choosing between two areas, both focused on saving children’s lives: malaria treatment in sub-Saharan Africa and childhood cancer treatment in the United States. Both are worthy causes, but an additional dollar spent on the first will likely save more lives in expectation than one spent on the second. Far more children—both in raw numbers and as a percentage of the population—die from malaria in sub-Saharan Africa than

<sup>3</sup>For more on Development Innovation Ventures at USAID, see <https://www.usaid.gov/div>.

<sup>4</sup>For more on Evidence Action’s Accelerator, see <https://www.evidenceaction.org/accelerator>.

<sup>5</sup>For more information on CHAI’s Incubator, see <https://www.clintonhealthaccess.org/news/givewell-funds-chai-led-new-program-incubator/>.

<sup>6</sup>For a list of open questions we are interested in, see <https://www.openphilanthropy.org/research/social-science-research-topics-for-global-health-and-wellbeing/>.

from cancer in the United States. Specifically, 387,000 people under the age of 20 died of malaria in sub-Saharan Africa in 2019 (a rate of 67.5 per 100,000); in contrast, roughly 7,420 people under the age of 20 died of neoplasms in high-income countries (a rate of 3.1 per 100,000), according to the 2019 Global Burden of Disease report (Global Burden of Disease Collaborative Network 2020). Moreover, seasonal malaria chemoprevention is cheaper than cancer treatment. Estimates across contexts in Burkina Faso, Chad, Mozambique, Nigeria, and Togo suggest that all cycles of seasonal malaria chemoprevention can be administered for an annual cost between \$5 and \$6.50 per child (GiveWell 2024).

Given these background facts, a grantmaker might begin by allocating funding to malaria treatment—focusing on the most cost-effective opportunities first, then moving to less cost-effective ones once those are exhausted—until the marginal dollar spent on malaria saves the same number of children’s lives in expectation compared to a dollar spent on cancer treatment. In this hypothetical example (no existing philanthropic organization is large enough to carry this example to its logical conclusion), the grantmaker may end up funding both malaria and cancer, but the malaria program would likely be much larger.

There are many objections to this style of analysis. For example, malaria chemoprevention is cheaper than cancer treatment, but cancer treatment might prevent certain death versus preventing a small chance of death. In addition, children who would have died of malaria might be more likely to be exposed to other risks and thus have shorter life expectancies than children who are cured of cancer. However, by sharpening our analysis further, we can address these concerns by estimating changes in the probability of survival, and by accounting for different life expectancies by considering years of life saved. Life expectancies do differ, but generally by much smaller margins than the costs of averting one death.

How do we define “benefits” across a wide range of grantmaking activities? How do we convert between health and economic benefits? How do we generate an overall measure of impact per dollar spent—such as cost-effectiveness or social return on investment—to evaluate opportunities and set a “bar” for grantmaking? The rest of this section walks through Open Philanthropy’s answers to these questions.

### **How Do We Define Economic Benefits across a Wide Range of Grantmaking Activities?**

To date, Open Philanthropy’s Global Health and Wellbeing work has largely focused on interventions that forestall premature death and improve morbidity (“health benefits”) and those that improve material deprivation (“economic benefits”). These are, of course, not the only outcomes that matter. Over time, we hope to consider a greater diversity of consequences. However, as a starting point, measures of health and economic benefits capture much of what is immediately important for the global poor, who are a major focus of our work.

To decide whether Open Philanthropy should launch a new philanthropic program in lead exposure instead of in a different cause area like climate change or tobacco control, we need a common currency for evaluating the potential economic

and health benefits that our grantmaking might accomplish in each of these areas. I will refer to these fungible units as “OP dollars.” My explanation here draws on Favaloro and Berger (2021).

To measure economic benefits, we use a logarithmic model of the utility of income—that is, a 1 percent increase in income is valued equally, regardless of the starting point. The log model offers simplicity and consistency with our reading of the literature (as a recent example, Elminejad, Havranek, and Irsova 2023). While we would prefer to capture consumption, we often use income as a stand-in because it is easier to measure. In addition, because the beneficiaries of our grantmaking activities are often people living in low- and middle-income countries, consumption and income generally do not differ greatly since savings and taxes are low.<sup>7</sup>

We define \$1 OP as the impact equivalent to raising the income of one person with an income of \$50,000 (which was approximately US GDP per capita when we first developed this framework) by \$1 for one year. In general, the value of an income change can be expressed as:

$$\$v = \$50,000 \times w \times \ln(1 + z\%) \times y.$$

Here, \$ $v$  OP corresponds to a  $z\%$  change in income for  $w$  number of people for  $y$  years. In other words, this economic intervention increases welfare by a multiple of  $v$  compared to giving \$1 to someone with an income of \$50,000.

Using this model, \$1 given to someone with an income of \$500 (that is, 100 times less than \$50,000) would be worth \$100 OP. Doubling a single person’s income for one year, regardless of starting point, is worth approximately \$35,000 OP ( $z = 100\%$ ,  $w = 1$ , and  $y = 1$ ):

$$\$v = \$50,000 \times 1 \times \ln(1 + 100\%) \times 1 = \sim \$35,000.$$

The idea that doubling someone’s income from \$50,000 to \$100,000 is worth \$35,000 OP arises because the logarithmic utility function encounters declining marginal returns over the course of a doubling; by the same logic, it is more valuable to increase 100 people’s incomes by 1 percent than one person’s income by 100 percent.

### **How Do We Define Health Benefits across a Wide Range of Grantmaking Activities?**

For health gains, we use “disability-adjusted life years,” often abbreviated as DALYs, a simplified metric for aggregating the severity and extent of different

<sup>7</sup>Data from Banerjee and Duflo (2009) shows under 15 percent of poor households around the world have a formal savings account. While informal mechanisms exist, less than 10 percent of poor households Banerjee and Duflo surveyed in India were part of an informal savings group. Households that do save are likely to be doing so as a precautionary measure to smooth consumption given variable incomes.

health conditions. The DALY total for a particular cause of death or risk factor is equal to the sum of “years of healthy life lost due to disability” and “years of life lost.”

One of the most widely used and comprehensive sources for global estimates of disability-adjusted life-years is the Global Burden of Disease (GBD).<sup>8</sup> The GBD is a systematic effort, led by the Institute for Health Metrics and Evaluation (IHME)<sup>9</sup> and its network of over 11,000 collaborators, to estimate and compare the burden of hundreds of causes of death and risk factors. It uses data from a variety of sources to assign all deaths and disability-adjusted life-years to an exhaustive list of causes in order to produce internally consistent and comprehensive estimates at the global, national, and subnational levels for a range of age and gender groups (for more on the evidence used to create these estimates, see Institute of Health Metrics and Evaluation 2020b, p. 23). The IHME publishes results in “rounds,” which provide estimates for new years, reestimates for all previous years going back to 1980, and forecasts to 2040 of future burdens.

The Global Burden of Disease defines “years lived with disability” by multiplying the duration of a condition by a “disability weight” that captures its severity (where 0 is full health and 1 is death), as judged via surveys of the general public. For example, chickenpox has a low disability weight (0.006) as it results in a low fever and mild discomfort, but does not cause major or lasting disruptions to daily life. In comparison, severe malaria has a higher disability weight (0.113) as it causes fever, pain, and fatigue which greatly disrupt life.<sup>10</sup>

The Global Burden of Disease defines “years of lives lost” by subtracting “the age at death from the longest possible life expectancy for a person at that age. For example, if the longest life expectancy for men in a given country is 75, but a man dies of cancer at 65, this would be 10 years of life lost due to cancer” (Institute for Health Metrics and Evaluation 2020a). One objection to this approach is that it does not take into account predicted life expectancy for a randomly selected individual with a particular condition. For instance, a person prevented from dying of one cause may face a greater risk of premature death than is suggested by the methodology above.

In short, a disability-adjusted life year for a given condition corresponds to years of healthy life lost to a cause of death or a risk factor, both in terms of shorter life

<sup>8</sup>The World Health Organization (WHO) similarly estimates burdens of disability-adjusted life-years for diseases in their Global Health Estimates (GHE), though it departs from the Global Burden of Disease in a few key ways. The GHE estimates tend to rely more on vital registration statistics provided by Member States, as well as modeling from WHO groups and UN agencies (WHO 2020a). Where these data do not exist, the GHE will draw upon the GBD’s estimates. The GHE also uses different assumptions; for example, by using the highest projected human life expectancy for 2050 rather than observed life expectancies (WHO 2020b). The WHO GHE tends to be less comprehensive than the GBD, making available for download only country-level burden estimates for the years 2000, 2005, 2010, and 2019 for a few age groups.

<sup>9</sup>The IHME has led the Global Burden of Disease since 2010. Previously, the GBD was led by the World Health Organization (Mathers 2020).

<sup>10</sup>For a full list of disability weights, see <https://ghdx.healthdata.org/record/ihme-data/gbd-2019-disability-weights>.

expectancy and partial years of healthy life lost according to the disability weights. There are many critiques of disability-adjusted life-years in general, and in particular of the use of disability weights. For example, these weights may not incorporate consequences unrelated to the health of the affected individual (say, grief experienced by loved ones), and sufficient consideration may not be given to the experiences of people living with the relevant conditions; for example, Karimi, Brazier, and Paisley (2017) find in a small survey that “the public [compared to patients] underestimated the effects of problems in usual activities compared to problems in mobility.”

We rely on disability-adjusted life years because measurements are available globally across a large number of health conditions via the Global Burden of Disease, though that does not mean we endorse all elements of the process used to elicit disability weights or specific values. A key advantage of the GBD is that it calculates DALYs across causes of death and risk factors in a consistent manner. One concern is that if we deviate from the GBD estimates for one condition because of a specific piece of evidence, we are uncertain whether we would deviate for all other conditions if we spent similar time evaluating them (though in reality we often try to corroborate the GBD estimates). Because we are ultimately undertaking a comparative exercise, consistent measurement is critical. We remain open to adopting alternative measures that allow for global comparisons, and we have looked into several options. However, we have yet to find one that meets our criteria for robustness and consistency. For the purposes of this paper, for example, I am taking the DALY estimate from the GBD for lead exposure as given, without adjustments. In practice, Open Philanthropy instead tries to estimate the counterfactual life expectancy of those affected by our work: if people avoid death from one cause, how long would they expect to live before dying for another reason? This change tends to decrease the health impact of many interventions we typically consider.

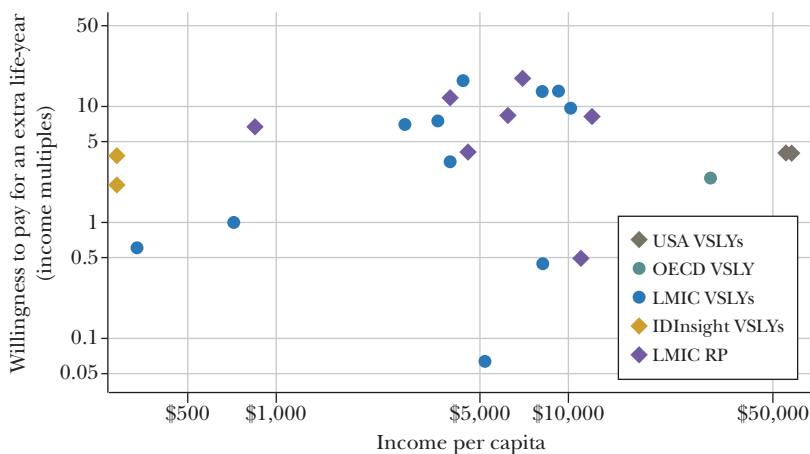
### **How Do We Convert between Health and Economic Benefits?**

We have set an exchange rate between health and economic benefits in order to allow us to compare interventions with different types and degrees of impact. We arrived at our current ratio on the basis of research that captures peoples’ revealed and stated preferences regarding this tradeoff.

Favaloro and Berger (2021) surveyed the “value of statistical life” literature.<sup>11</sup> Figure 1, reproduced from that blog post, plots the estimates found in Robinson, Hammitt, and O’Keefe’s (2019) meta-analysis. Some studies (shown as diamonds) use actual choices that workers make between jobs with different wages and health risks, or choices that consumers make in purchasing health-protecting items, to reveal the value placed on reducing the risk of lost life. Other studies (shown as circles) use survey tools that seek to elicit a stated value of reducing risks to life compared to other values. The two yellow diamonds represent findings from a paper

<sup>11</sup> The concept of the “value of a statistical life” is controversial. Objections include that it is wrong to put any monetary value on life and that relying on stated and revealed preferences for the “value of a statistical life” can imply unacceptable variation (for example, across income levels), among other concerns.

*Figure 1*  
**Willingness to Pay for an Extra Life-Year as a Multiple of Annual Income per Capita**



*Source:* Reproduced from Favaloro and Berger (2021).

*Note:* This figure plots estimates found in Robinson, Hammitt, and O’Keefe’s (2019) meta-analysis, expressed as willingness to pay for an extra life-year as a multiple of annual income per capita. VSLY stands for “value of a statistical life-year.” LMIC RP stands for studies that show revealed preference in low- and middle-income countries.

conducted by Redfern et al. (2019), commissioned by GiveWell, which surveyed people demographically similar to the beneficiaries of GiveWell’s recommended charities. Most of these studies report a willingness to pay between 0.5 and 10 times annual income for an extra year of healthy life.

We have also considered the subjective wellbeing literature, which asks individuals to rate life satisfaction, typically on a scale from one to ten points. One needs a model to translate reported life satisfaction into subjective wellbeing and to imply a tradeoff with mortality, but it is possible to solve for the change in income that would generate a similar increase in satisfaction-point-years as extending life by one year. In general, relying more heavily on this evidence would lead to a greater emphasis on health (as opposed to economic) benefits than our current approach.

On the basis of this research, we decided provisionally to value a disability-adjusted life-year approximately three times more than doubling a person’s income for one year, as this lies around in the middle of the research findings surveyed above. This multiple is also roughly consistent with guidance from other actors focused on global health; for example, the World Health Organization previously recommended a cost-effectiveness threshold of three times per capita GDP (Griffiths, Maruszczak, and Kusel 2015).

Our valuation of a disability-adjusted life year is \$100,000 in the metric of OP dollars described above, or approximately three times \$35,000 OP (the value of



an income doubling for a single individual for one year described above). Notably, the valuation of \$100,000 OP cannot be directly compared to local incomes because we value absolute dollars very differently across contexts (according to our logarithmic utility function; for example, \$1 given to someone living on \$500 per year is worth \$100 OP).

We think our valuation on health relative to economic benefits may be too low based on the evidence in Figure 1, but most of Open Philanthropy’s Global Health and Wellbeing work already focuses on health, so we view the decision not to place an even higher valuation on health as conservative. Placing a value on this tradeoff is a consequential decision for our prioritization of projects and remains a work in progress. We hope to find more projects like Redfern et al. (2019) that generate new evidence on this question, focused on people living in low- and middle-income countries.

### **How Do We Generate an Overall Measure of Impact per Dollar Spent to Evaluate Opportunities and Set a “Bar” for Grantmaking?**

We compare individual opportunities based on their social return on investment, equal to expected benefits expressed in OP dollars divided by costs in absolute (that is, not OP) dollars. We define a unit of cost-effectiveness as the return on investment generated by giving \$1 to someone with an income of \$50,000, which is equivalent to extending an individual’s life by one healthy year at a cost of \$100,000.

To set a cost-effectiveness bar, we try to estimate the philanthropic return on investment of our least cost-effective dollar—what we hope is our last dollar if we correctly prioritize the most cost-effective opportunities first. On a day-to-day basis, we decide to make a grant if and only if we believe it will—in expectation—generate more health and economic benefits per dollar than our bar. If not, we save that funding in the hopes of finding another, better opportunity.<sup>12</sup> If we consistently cannot find interventions above our bar, we incrementally lower it. In 2021, we published details describing how we settled on a bar of a cost-effective return of 1,000 (Favorolo and Berger 2021). Since then, we have continued to update this value based on our best understanding of available philanthropic opportunities.

### **Importance, Neglectedness, and Tractability**

“Equalizing marginal returns” is a useful framework for evaluating and comparing specific interventions, but it often does not produce practical recommendations when considering unfamiliar problem areas where we do not yet know

<sup>12</sup>We have also considered how to approach intertemporal optimization. Drawing on work from Trammell (2021) and others, we have developed a model to weigh factors incentivizing spending sooner (for example, cost effective opportunities may become harder to find as the world improves or the possibility of high-impact philanthropy growing over time) and spending later (for example, taking advantage of market returns in the meantime before acting, and more time to learn).

what interventions are feasible. To evaluate cause areas at a high level, we use the INT framework.<sup>13</sup> For importance, how many are affected by the problem and how intensely? For neglectedness, how much attention does this problem currently receive and how much is it likely to receive in the future, absent intervention from Open Philanthropy? For tractability, can a philanthropic funder make headway?

Here, I apply the INT framework to the issue of lead exposure. I also discuss the heuristics to complement the INT framework, along with an example of how this has led us to consider grantmaking in the area of farm animal welfare.

### **Applying the INT Framework to Lead Exposure**

The 2019 Global Burden of Disease results—the most recent figures available publicly in full—estimate that lead exposure accounts for approximately 900,000 deaths and approximately 22 million disability-adjusted life-years lost annually by exacerbating risk factors for cardiovascular diseases, mental disorders, and kidney disease (Global Burden of Disease Collaborative Network 2020). Over 95 percent of the burden is concentrated outside of high-income countries. To provide a rough comparison with other causes of death, lead exposure would fall between HIV/AIDS (approximately 864,000 annual deaths) and tuberculosis (approximately 1.18 million annual deaths). The 2021 GBD results—which are only partially available—estimate that lead exposure is responsible for 1.57 million annual deaths (Global Burden of Disease Collaborative Network 2023). The World Bank offers an even higher figure, considering mediation through mechanisms in addition to hypertension, estimating 5.5 million annual deaths attributable to lead (Larsen and Sánchez-Triana (2023)).

Lead exposure is commonly quantified via the concentration of lead in the blood and measured in units of micrograms per deciliter ( $\mu\text{g}/\text{dL}$ ). In a systematic review of studies published between 2010 and 2019 reporting blood lead levels in low- and middle-income countries, Ericson et al. (2021) found that among countries with sufficiently representative data, 49 percent of children had blood lead levels over 5 micrograms per deciliter.<sup>14</sup> For reference, in Flint, Michigan, the center of a major scandal on child lead exposure in the United States, the estimated percentage of children above  $5\mu\text{g}/\text{dL}$  was approximately 12 percent in 2006 and 3 percent in 2016 (Gómez et al. 2018).

A wide array of evidence supports a connection between lead exposure and poor health, including animal evidence (for example, as summarized in US EPA 2013), epidemiological evidence based on correlations between lead exposure and health and economic effects (for meta-analyses, see Navas-Acien et al. 2007; Lanphear et al. 2018), and quasi-experiments, such as when NASCAR shifted its car-racing

<sup>13</sup>We describe the INT framework in more detail at <https://www.openphilanthropy.org/cause-selection/>.

<sup>14</sup>According to the Center for Global Development’s “A Call to Action to End Childhood Lead Poisoning Worldwide”: “The  $5\mu\text{g}/\text{dL}$  is not a biological threshold of great significance, but represents a standard formerly used by the US Centers for Disease Control to indicate higher lead exposure than most US children; it is still used by the WHO and commonly applied as a benchmark reference level” (<https://www.cgdev.org/sites/default/files/call-action-end-childhood-lead-poisoning-worldwide.pdf>).

to unleaded gas in 2007 (Hollingsworth and Rudik 2021), or when regulations caused US battery-recycling sites to shift to Mexico in 2009 (Tanaka, Teshima, and Verhoogen 2022).

Indeed, although we rely on the Global Burden of Disease estimates for purposes of comparability, our best guess is that this report has underestimated lead's health burden. This is in part due to modeling choices that overestimate the "safe" level of lead exposure (Shaffer et al. 2019) and in part due to health outcomes not included in the GBD study, such as stillbirths, which may increase with lead exposure (Clay, Hollingsworth, and Severnini 2023). In addition, lead exposure also has cognitive effects. The research literature has substantiated a biological mechanism by which lead can impair cognitive ability by replacing calcium in the brain (Rădulescu and Lundgren 2019), and includes both well-identified animal studies of lead affecting cognition (Anderson, Mettil, and Schneider 2016; Rice and Gilbert 1985) and several meta-analyses summarizing observational studies and natural experiments in humans that consistently find a negative effect of lead exposure on IQ (Larsen and Sánchez-Triana 2023; Crawford et al. 2023; Schwartz 1994; Lanphear et al. 2005).

Estimating the magnitude of lead exposure on economic outcomes is difficult. Childhood lead exposure and measures of parental socioeconomic status appear to be correlated, and given that most evidence in humans is observational, it is difficult to know whether the effect sizes are overstated because of persistent omitted variables (for one attempt to unpack these issues, see Lanphear et al. 2005, especially Table 4). Still, taken as a whole, we find the collective evidence indicating a link between lead exposure and cognitive outcomes compelling.

To evaluate neglectedness, we try to estimate the total amount of "relevant" spending on a given issue. "Relevant" is difficult to define crisply, but it broadly refers to funding that targets the same problem through channels we might also consider funding (that is, we typically exclude private investments as well as spending on basic research elucidating the mechanisms of lead exposure's adverse impacts). Several years ago, collective efforts to combat lead exposure in low- and middle-income countries received less than \$10 million annually in philanthropic funding (Bernard and Schukraft 2021), and in 2023 that number increased to just over \$11 million (Bonnifield and Todd 2023). We estimate that there is currently approximately \$10–15 million per year in relevant spending.<sup>15</sup> In general, when total estimates are not readily available, we are often able to use public data sources and conversations with peers at other funders to calculate a rough order-of-magnitude sense of the degree of "relevant" funding devoted to some issue.

In determining tractability, despite the limited philanthropic attention focused on lead exposure so far, some interventions suggest that the issue could be addressed

<sup>15</sup>The main organizations working to reduce lead exposure include International Pollutants Elimination Network and its member organizations, Lead Exposure Elimination Project, Global Alliance to Eliminate Lead Paint, Global Alliance on Health and Pollution, Occupational Knowledge International, Pure Earth, Vital Strategies, the Center for Global Development, and UNICEF. Key funders supporting this work include GiveWell, USAID, the Swedish International Development Agency, the Global Environment Facility, the Clarios Foundation, and the World Bank.

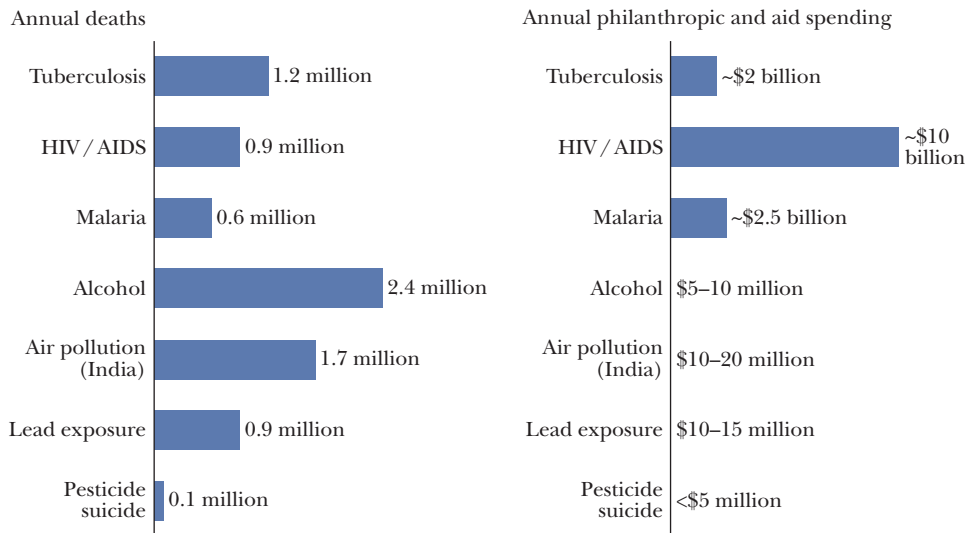
inexpensively. For example, until recently, a major source of lead poisoning in Bangladesh came from the use of lead chromate in brightening turmeric (a commonly-used cooking spice), identified by a group of researchers at Stanford University and a Bangladeshi nonprofit organization called icddr.<sup>b</sup> By testing samples in markets, advocating for compliance among suppliers, and enacting government fines to enforce existing policy, the research team and the government of Bangladesh were successful in decreasing contamination dramatically, lowering the proportion of contaminated turmeric samples from 47 percent in 2019 to 0 percent in 2021 (Forsyth et al. 2023). Similarly, Pure Earth has partnered with the government of the Republic of Georgia to identify sources of lead in spices and reduce contamination via new regulations, consumer and producer education, heightened monitoring of spices at markets, and increased regulatory enforcement (Berg 2022). In addition, Lead Exposure Elimination Project (LEEP) has had success working with the governments of Malawi (LEEP 2021) and Pakistan (LEEP 2023) to implement lead paint regulation. Tractability is typically the hardest criterion for us to assess when exploring new causes, and we find that many potential causes fall in a vague, middling range, making them difficult to differentiate along this dimension.

### **Heuristics to Complement the INT Framework**

Within our INT framework, we often rely on a few additional heuristics. One such heuristic compares the ratio of neglectedness to importance across different problems. To give an example, for lead exposure, our best guess of philanthropic spending in reducing lead exposure per disability-adjusted life-year lost is 45 cents, which is two orders of magnitude lower than our estimate of the equivalent figure for malaria and tuberculosis, themselves neglected diseases, and at least three orders of magnitude lower than our estimate for cancer (see Figure 2 for some of these comparisons). To present another similar metric, lead exposure constitutes approximately 1 percent of the global disease burden; using the estimate of \$150 billion in annual philanthropic funding (Johnson 2018), the \$10 million in annual funding directed at lead constitutes less than 0.01 percent, again a difference exceeding two orders of magnitude. Though the comparison does not reveal anything about tractability—whether it is more difficult to make progress in lead, malaria, and tuberculosis compared to cancer—it is a useful rule of thumb that suggests there may be overlooked low-hanging fruit. In this instance, these areas are likely neglected at least in part because the victims of lead, malaria, and tuberculosis are concentrated in low- and middle-income countries instead of high-income ones. In other cases, opportunities might be neglected because they are relatively new, technically or logistically complex, and/or high-risk.

A second heuristic is to look for opportunities that present a special role for philanthropy relative to market-based or democratic mechanisms (this paragraph paraphrases an idea described internally by Open Philanthropy's CEO, Alexander

Figure 2

**Annual Deaths Compared to Annual Philanthropic and Aid Spending**

Source: Reproduced from Snowden (2023).

Note: This figure compares annual deaths with annual philanthropic and aid spending attributable to various diseases and risk factors. “Pesticide suicide” refers to self-poisoning with agricultural pesticides.

Berger).<sup>16</sup> In many of the cause areas where we believe our support has been most influential, the beneficiaries are often structurally underrepresented. For example, we support advocacy to reduce barriers to immigration, especially from low-income to high-income countries, and to combat restrictive local land use regulations. In both areas, political systems—which tend to focus on the interests of existing constituents—can undervalue the interests of potential participants who would benefit from more liberal policies. We also allocate a large portion of our giving to global health and development challenges specific to low-income countries. These problems almost always receive fewer resources than comparable health challenges in high-income countries. Most inequality occurs between rather than within countries, and democratic institutions in high-income states typically have limited interest in addressing this discrepancy.

A third heuristic that we use for neglectedness considers expanding one’s moral circle to include those not widely recognized as in need of “empathy or moral concern” (Karnofsky 2017). Our work on farm animal welfare is most representative of this lens: most grantmaking organizations do not consider animals as

<sup>16</sup>For links at the Open Philanthropy website to topics mentioned in this paragraph, see <https://www.openphilanthropy.org/focus/immigration-policy/> for immigration policy, <https://www.openphilanthropy.org/focus/land-use-reform/> for land use reform, and <https://www.openphilanthropy.org/focus/global-health-development/> for global health and development.

targets of philanthropic attention. As a result, we see outsized opportunities for impact, as some “quick-win” grants in this area would otherwise not be made. That said, we try to avoid prioritizing novelty as an end in itself. We aspire to “radical empathy,” but only in the sense that “our goal is to do the most good we can, not to seek out and support those causes which are most ‘radical’ in our present society” (Karnofsky 2017).

### **The Outlier Opportunities Principle**

Areas that are important, neglected, and tractable seem more likely to offer highly cost-effective, specific interventions upon further investigation. This approach can also capture what we call the “outlier opportunities principle”: “If we see an opportunity to do a huge, and in some sense ‘unusual’ or ‘outlier’ amount of good according to worldview A by sacrificing a relatively modest, and in some sense ‘common’ or ‘normal’ amount of good according to worldview B, we should do so (presuming that we consider both worldview A and worldview B highly plausible and reasonable and have deep uncertainty between them)” (Karnofsky 2018).

Open Philanthropy applies this principle in contexts beyond our Global Health and Wellbeing work. For example, via our farm animal welfare grantmaking, we spend millions of dollars each year supporting efforts to improve the lives of animals confined on factory farms.<sup>17</sup> It is extremely difficult to compare human lives lost due to, say, tuberculosis to the experience of billions of chickens confined to battery cages. That said, the cause of farm animal welfare stands out under the INT framework. By one estimate, US animal shelters and rescue groups (often for dogs and cats) raise approximately \$2.8 billion per year (IBISWorld 2023),<sup>18</sup> or approximately \$431 for each of the 6.5 million animals that enter shelters and rescues annually (ASPCA 2024). By contrast, farm animal advocacy groups raise approximately \$90 million per year for work in the United States (Farmed Animal Funders 2021), which equates to approximately \$0.01 for each of the 10 billion animals farmed last year (or around \$0.04 for each of the 2.5 billion farm animals alive at any time) (Ritchie, Rosado, and Roser 2023). We believe that before Open Philanthropy began grantmaking in this area, the space was at least relatively neglected. Finally, tractability seems high, because there are clear ways for a philanthropist to make progress, for instance by supporting advocacy to secure corporate commitments related to animal welfare, or by investing in research into animal product alternatives.<sup>19</sup> Even if animals’ capacity to experience suffering is only a

<sup>17</sup>For more on our farm animal welfare grantmaking, see <https://www.openphilanthropy.org/focus/farm-animal-welfare/>.

<sup>18</sup>Based on conversations with experts, we believe this number is an underestimate. The actual figure is likely closer to \$4 billion a year, which would be approximately \$615 spent for each shelter animal.

<sup>19</sup>For an analysis of funding in farm animal welfare before we began grantmaking, see <https://www.openphilanthropy.org/research/treatment-of-animals-in-industrial-agriculture/#3-who-else-is-working-on-this>. For a few “wins” in our farm animal welfare portfolio, which underscores the area’s tractability, see <https://www.openphilanthropy.org/research/a-year-of-wins-for-farmed-animals/>.

small fraction of humans', funding in this area may alleviate much more suffering than other opportunities due to its sheer scale.

## Critiques and Limitations

Comparing philanthropic cause areas is hard, and in practice allocation decisions are much messier than the framing I have used above may imply.

Our frameworks are full of uncertainties, and at times we may not implement it well. To give a few examples within the issue of reducing lead exposure, many of the underlying pieces of evidence have small sample sizes, are potentially vulnerable to confounding variables, and do not cover the populations we aim to target with our grantmaking. These issues could lead to large errors in our estimates of the size of the problem. We also do not have a robust sense of how blood lead levels relate to different sources of lead exposure, and so when moving to practical grantmaking, we might not address the most important sources of contamination. We could also be undercounting harms from reducing lead in products. Additionally, we might be underestimating the extent to which future health technologies (for example, better drugs for safely chelating lead out of blood) will address the harms posed by lead, which would reduce the counterfactual impact of taking action today.

In addition, our framework may inherently produce important blind spots. We may be prone to overrate marginal changes and underrate the possibility for transformation. For example, our framework might overlook basic science opportunities with unknown paths to impact. Although we conduct surveys in an attempt to capture beneficiary preferences,<sup>20</sup> we are almost certainly missing decision-relevant information. We strive for consistency, but we often struggle to capture systematically the many different benefits and harms a single intervention can generate, especially when those effects extend far into the future and/or are indirect. It is also difficult to trace the causal effect of our funding on an organization's behavior, and of that behavior's impact on the world.

We know we are not the first to encounter these dilemmas, and we do not believe we have perfected implementation. Any attempt to define the benefits of philanthropic work touches on long-standing philosophical and economic questions, and we recognize that we are frequently less-than-fully-correct or just outright wrong. We aim to keep all of this in mind, and approach our work with humility.

That said, we also believe that setting clear goals and defining acceptable tradeoffs have forced us to confront potential biases when determining what causes deserve our empathy and support. Further, this framework has provided us with the ability to compare causes on a relatively equal footing with rigor and transparent assumptions. As we continue to improve our work and our understanding of the world—including as beneficiaries of a growing body of social scientific

<sup>20</sup>For example, see <https://www.openphilanthropy.org/grants/idinsight-beneficiary-preferences-survey/>.

knowledge—we think that this approach will lead us to support opportunities with greater positive impact per dollar than we would otherwise find.

■ *This paper is based on research and grantmaking at Open Philanthropy and GiveWell. Thank you to the full staff (current and past) of both organizations. Thank you to Norma Altshuler, Alexander Berger, Lewis Bollard, Jasmine Dhaliwal, Peter Favaloro, Graham Gottlieb, Paige Henchen, Teryn Mattox, Jason Schukraft, and Jacob Trefethen for helpful comments. Special thanks to Britney Budimen, Aaron Gertler, Otis Reid, Chris Smith, and James Snowden for their detailed feedback. Thank you to Jamie Simonson for excellent research assistance. Thank you to Rethink Priorities for early research on lead exposure and Rachel Silverman Bonnifield (Center for Global Development), Lucia Coulter (Lead Exposure Elimination Project), Clare Donaldson (Lead Exposure Elimination Project), Jenna Forsyth (Stanford University), Bruce Lanphear (Simon Fraser University), Drew McCartor (Pure Earth), Rich Fuller (Pure Earth), and many others, who are making real progress on this issue. Finally, thank you to Nina Pavcnik, Timothy Taylor, and Heidi Williams from the JEP team for the thoughtful feedback and edits. Any errors are mine alone.*

## References

- Albrecht, Björn, Sandra Andersen, Keval Chauhan, and Daina Graybosch, and Philippe Menu.** 2018. “Pursuing Breakthrough in Cancer Drug Development.” McKinsey and Company, January 12. <https://www.mckinsey.com/industries/life-sciences/our-insights/pursuing-breakthroughs-in-cancer-drug-development>.
- Anderson, D. W., W. Mettil, and J. S. Schneider.** 2016. “Effects of Low Level Lead Exposure on Associative Learning and Memory in the Rat: Influences of Sex and Developmental Timing of Exposure.” *Toxicology Letters* 246: 57–64.
- Anthi, Kelly, and Jacy Reese Anthi.** 2019. “Global Farmed and Factory Farmed Animals Estimates.” Sentience Institute, February 21. <https://www.sentienceinstitute.org/global-animal-farming-estimates>.
- ASPCA.** 2024. “Pet Statistics.” American Society for the Prevention of Cruelty to Animals. <https://www.aspc.org/helping-people-pets/shelter-intake-and-surrender/pet-statistics>.
- Banerjee, Abhijit V., and Esther Duflo.** 2007. “The Economic Lives of the Poor.” *Journal of Economic Perspectives* 21 (1): 141–61.
- Berg, Sarah.** 2022. “Republic of Georgia: Solving the Mystery of Widespread Lead Poisoning in Georgian Children.” Pure Earth, May 3. <https://www.pureearth.org/solving-the-mystery-of-widespread-lead-poisoning-in-georgian-children>.
- Bernard, David Rhys, and Jason Schukraft.** 2021. “Global Lead Exposure Report.” Rethink Priorities, May 29. <https://rethinkpriorities.org/publications/global-lead-exposure-report>.
- Bonnifield, Rachel Silverman, and Rory Todd.** 2023. “A Call to Action to End Lead Poisoning Worldwide.” Washington, DC: Center for Global Development.
- Centers for Disease Control.** 2022. “Blood Lead Reference Value.” CDC, December 2. <https://www.cdc.gov/nceh/lead/data/blood-lead-reference-value.htm>.



- Clay, Karen, Alex Hollingsworth, and Edson R. Severnini.** 2023. “The Impact of Lead Exposure on Fertility, Infant Mortality, and Infant Birth Outcomes.” NBER Working Paper 31379.
- Crawford, Lee, Rory Todd, Susannah Hares, Justin Sandefur, and Rachel Silverman Bonnifield.** 2023. “How Much Would Reducing Lead Exposure Improve Children’s Learning in the Developing World?” Center for Global Development Working Paper 650.
- Elminejad, Ali, Tomas Havranek, and Zuzana Irsova.** 2023. “Estimating Relative Risk Aversion from the Euler Equation: The Importance of Study Design and Publication Bias.” <https://elminejad.github.io/Website/RRA.pdf>.
- Ericson, Bret, Howard Hu, Emily Nash, and Greg Ferraro, Julia Sinitsky, and Mark Patrick Taylor.** 2021. “Blood Lead Levels in Low-Income and Middle-Income Countries: A Systematic Review.” *The Lancet: Planetary Health* 5 (3): E145–53.
- Farmed Animal Funders.** 2021. “State of the Movement Report: 2021.” <https://static1.squarespace.com/static/645be31efb5e013b9c6f4751/t/64cb9a7bb5b56917db63457a/1691064968896/FAF+State+of+the+Movement+Report+2021.pdf>.
- Favaloro, Peter, and Alexander Berger.** 2021. “Technical Updates to Our Global Health and Well-being Cause Prioritization Framework.” Open Philanthropy, November 18. <https://www.openphilanthropy.org/research/technical-updates-to-our-global-health-and-wellbeing-cause-prioritization-framework/>.
- Forsyth, Jenna E., Musa Baker, Syeda Nurunnahar, Shariful Islam, M. Saiful Islam, Tauhidul Islam, Erica Plambeck et al.** 2023. “Food Safety Policy Enforcement and Associated Actions Reduce Lead Chromate Adulteration in Turmeric across Bangladesh.” *Environmental Research* 232: 116328.
- GiveWell.** 2024. “Seasonal Malaria Chemoprevention.” <https://www.givewell.org/international/technical/programs/seasonal-malaria-chemoprevention>.
- Global Burden of Disease Collaborative Network.** 2020. “Global Burden of Disease Study 2019 (GBD 2019).” Institute for Health Metrics and Evaluation (IHME). <https://www.healthdata.org/research-analysis/gbd>.
- Global Burden of Disease Collaborative Network.** 2023. “Global Burden of Disease Study 2021 (Preview GBD Update).” Institute for Health Metrics and Evaluation (IHME).
- Gómez, Hernán F., Dominic A. Borgianni, Mahesh Sharman, Keneil K. Shah, Anthony J. Scolpino, James M. Oleske, and John D. Bogden.** 2018. “Blood Lead Levels of Children in Flint, Michigan: 2006–2016.” *Journal of Pediatrics* 197: 158–64.
- Griffiths, M., M. Maruszczak, and J. Kusel.** 2015. “The WHO-Choice Cost-Effectiveness Threshold: A Country-Level Analysis of Changes over Time.” *Value in Health* 18 (3): PA88.
- Hollingsworth, Alex, and Ivan Rudik.** 2021. “The Effect of Leaded Gasoline on Elderly Mortality: Evidence from Regulatory Exemptions.” *American Economic Journal: Economic Policy* 13 (3): 345–73.
- IBISWorld.** 2023. “Animal Rescue Shelters in the US—Market Size, Industry Analysis, Trends and Forecasts (2024–2029).” IBISWorld Industry Reports, May. <https://www.ibisworld.com/united-states/market-research-reports/animal-rescue-shelters-industry/#IndustryStatisticsAndTrends>.
- Institute of Health Metrics and Evaluation.** 2020a. “Global Burden of Disease (GBD): About GBD.” Institute of Health Metrics and Evaluation: Research and analysis. <https://www.healthdata.org/research-analysis/about-gbd>.
- Institute of Health Metrics and Evaluation.** 2020b. “Protocol for the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD).” Institute of Health Metrics and Evaluation. [https://www.healthdata.org/sites/default/files/files/Projects/GBD/March2020\\_GBD%20Protocol\\_v4.pdf](https://www.healthdata.org/sites/default/files/files/Projects/GBD/March2020_GBD%20Protocol_v4.pdf).
- Johnson, Paula.** 2018. *Global Philanthropy Report: Perspectives on the Global Foundation Sector*. Cambridge, MA: Harvard Kennedy School.
- Karimi, M., J. Brazier, and S. Paisley.** 2017. “Are Preferences over Health States Informed?” *Health and Quality of Life Outcomes* 15: 105.
- Karnofsky, Holden.** 2017. “Radical Empathy.” Open Philanthropy, February 16. <https://www.openphilanthropy.org/research/radical-empathy>.
- Karnofsky, Holden.** 2018. “Update on Cause Prioritization at Open Philanthropy.” Open Philanthropy, January 26. <https://www.openphilanthropy.org/research/update-on-cause-prioritization-at-openphilanthropy>.
- Kremer, Michael, Stephen P. Luby, Ricardo Maertens, Brandon Tan, and Witold Więcek.** 2023. “Water Treatment and Child Mortality: A Meta-Analysis and Cost-Effectiveness Analysis.” NBER Working 30835.

- Larsen, Bjorn, and Ernesto Sánchez-Triana.** 2023. "Global Health Burden and Cost of Lead Exposure in Children and Adults: A Health Impact and Economic Modelling Analysis." *The Lancet Planetary Health* 7 (10): E831–40.
- Lajoie, Julie.** 2015. "Understanding the Measurement of Global Burden of Disease." National Collaborating Centre for Infectious Diseases. <https://nccid.ca/publications/understanding-the-measurement-of-global-burden-of-disease>.
- Lanphear, Bruce P., Richard Hornung, Jane Khoury, Kimberly Yolton, Peter Baghurst, David C. Bellinger, Richard L. Canfield et al.** 2005. "Low-Level Environmental Lead Exposure and Children's Intellectual Function: An International Pooled Analysis." *Environmental Health Perspectives* 113 (7): 894–99.
- Lanphear, Bruce P., Stephen Rauch, Peggy Auinger, Ryan W. Allen, and Richard W. Hornung.** 2018. "Low-Level Lead Exposure and Mortality in US Adults: A Population-Based Cohort Study." *The Lancet Public Health* 3 (4): E177–84.
- Lead Exposure Elimination Project (LEEP).** 2021. "LEEP's First Year in Review." LEEP, December 29. <https://leadelimination.org/annual-review-2020-21>.
- LEEP.** 2023. "Progress on Lead Paint in Pakistan." LEEP, July 4. <https://leadelimination.org/progress-on-lead-paint-in-pakistan>.
- Mathers, Colin D.** 2020. "History of Global Burden of Disease Assessment at the World Health Organization." *Archives of Public Health* 78: 77.
- Murray, Christopher J. L.** 2022. "The Global Burden of Disease Study at 30 Years." *Nature Medicine* 28: 2019–26.
- Navas-Acien Ana, Eliseo Guallar, Ellen K. Silbergeld, and Stephen J. Rothenberg.** 2007. "Lead Exposure and Cardiovascular Disease—A Systematic Review." *Environmental Health Perspectives* 115 (3): 472–82.
- Philipson, Tomas J., Yier Lang, and Ruiquan Chang.** 2022. "The Impact of Recent White House Proposals on Cancer Research." University of Chicago Policy Brief. <https://ecchc.economics.uchicago.edu/project/policy-brief-the-impact-of-recent-white-house-proposals-on-cancer-research>.
- Rădulescu, Anca, and Stephen Lundgren.** 2019. "A Pharmacokinetic Model of Lead Absorption and Calcium Competitive Dynamics." *Scientific Reports* 9: 14225.
- Redfern, Alice, Martin Gould, Maryanne Chege, Sindy Li, and William Slotznick.** 2019. "Beneficiary Preferences Findings from Ghana and Kenya." November 7. [https://files.givewell.org/files/DWDA%202009/IDinsight/IDinsight\\_Beneficiary\\_Preferences\\_Final\\_Report\\_November\\_2019.pdf](https://files.givewell.org/files/DWDA%202009/IDinsight/IDinsight_Beneficiary_Preferences_Final_Report_November_2019.pdf).
- Rice, Deborah C., and Steven G. Gilbert.** 1985. "Low Lead Exposure from Birth Produces Behavioral Toxicity (DRL) in Monkeys." *Toxicology and Applied Pharmacology* 80 (3): 421–26.
- Ritchie, Hannah, Pablo Rosado, and Max Roser.** 2023. "Meat and Dairy Production." Our World in Data, December. <https://ourworldindata.org/meat-production>.
- Robinson, Lisa A., James K. Hammitt, and Lucy O'Keefe.** 2019. "Valuing Mortality Risk Reductions in Global Benefit-Cost Analysis." *Journal of Benefit-Cost Analysis* 10 (S1): 15–50.
- Roodman, David.** 2023. "Schooling and Labor Market Consequences of School Construction in Indonesia: Comment." Institute for Replication Discussion Paper Series 15.
- Schwartz, Joel.** 1994. "Low-Level Lead Exposure and Children's IQ: A Metaanalysis and Search for a Threshold." *Environmental Research* 65 (1): 42–55.
- Shaffer, Rachel M., Samuel P. Sellers, Marissa G. Baker, Rebeca de Buen Kalman, Joseph Frostad, Megan K. Suter, Susan C. Anenberg et al.** 2019. "Improving and Expanding Estimates of the Global Burden of Disease Due to Environmental Health Risk Factors." *Environmental Health Perspectives* 127 (10): 105001.
- Snowden, James.** 2023. "New Grantmaking Program: Global Public Health Policy." Open Philanthropy, November 22. <https://www.openphilanthropy.org/research/new-grantmaking-program-global-public-health-policy>.
- Stevenson, Betsey, and Justin Wolfers.** 2013. "Subjective Well-Being and Income: Is There Any Evidence of Satiation." *American Economic Review* 103 (3): 598–604.
- Tanaka, Shinsuke, Kensuke Teshima, and Eric Verhoogen.** 2022. "North-South Displacement Effects of Environmental Regulation: The Case of Battery Recycling." *American Economic Review: Insights* 4 (3): 271–88.
- Trammell, Philip.** 2021. "Dynamic Public Good Provision under Time Preference Heterogeneity: Theory and Applications to Philanthropy." Global Priorities Institute Working Paper 9-2021.

- US EPA.** 2013. *Integrated Science Assessment (ISA) for Lead (Final Report, Jul 2013)*. EPA/600/R-10/075F. Washington, DC: US Environmental Protection Agency.
- WHO.** 2020a. "WHO Methods and Data Sources for Country-Level Causes of Death 2000–2019." Global Health Estimates Technical Paper WHO/DDI/DNA/GHE/2020.2.
- WHO.** 2020b. "WHO Methods and Data Sources for Global Burden of Disease Estimates 2000–2019." Global Health Estimates Technical Paper WHO/DDI/DNA/GHE/2020.3.



# The Shifting Reasons for Beveridge Curve Shifts

Gadi Barlevy, R. Jason Faberman, Bart Hobijn,  
and Ayşegül Şahin

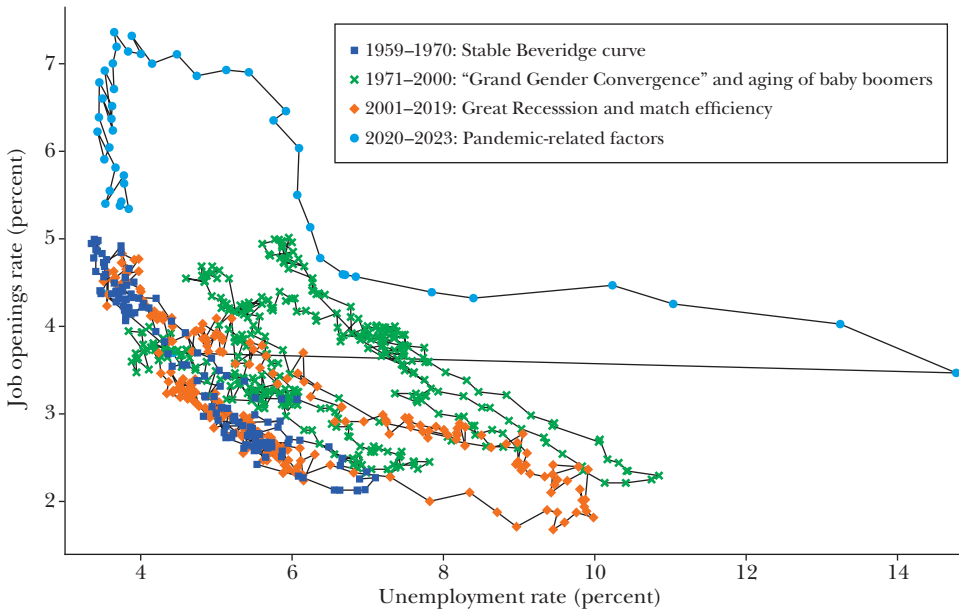
**I**n his groundbreaking analysis, William Beveridge (1944) argued that most fluctuations in unemployment are driven by changes in the demand for workers and that job openings are a useful measure of this demand. This insight implies a negative relationship between job openings and the unemployment rate. Although Beveridge never plotted the data, graphs showing this relationship—as plotted in Figure 1 for the US economy for 1959 through 2023—bear his name. On the figure's horizontal axis is the unemployment rate. On the vertical axis is the job openings rate based on data from the Job Openings and Labor Turnover Survey (JOLTS), which began in 2000, along with historical data on job openings derived by Petrosky-Nadeau and Zhang (2021) for 1959 to 2000.<sup>1</sup>

■ *Gadi Barlevy is a Senior Economist and Economic Adviser, Federal Reserve Bank of Chicago, Chicago, Illinois, and a Research Fellow, Institute of Labor Economics (IZA), Bonn, Germany. R. Jason Faberman is a Senior Economist and Economic Adviser, Federal Reserve Bank of Chicago, Chicago, Illinois, and a Research Fellow, Institute of Labor Economics (IZA), Bonn, Germany. Bart Hobijn is a Senior Economist and Economic Adviser, Federal Reserve Bank of Chicago, Chicago, Illinois. Ayşegül Şahin is a Professor of Economics, University of Texas at Austin, Austin, Texas; Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts; and a Research Fellow, Institute of Labor Economics (IZA), Bonn, Germany. Their email addresses are gbarlevy@frbchi.org, jfaberman@frbchi.org, bart.hobijn@barthobijn.net, and aysegul.sahin@austin.utexas.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.2.83>.

<sup>1</sup>The Job Openings and Labor Turnover Survey (JOLTS) is a monthly survey of about 21,000 establishments, conducted by the Bureau of Labor Statistics (see <https://www.bls.gov/jlt/>). It provides data on job openings, hires, layoffs, quits, and other separations. The job openings estimates derived by Petrosky-Nadeau and Zhang (2021) are an econometric splice of estimates from Barnichon (2010) and the Help Wanted Index, previously produced by the Conference Board. The job openings series

Figure 1

**US Beveridge Curve: January 1959–December 2023**

Source: US Bureau of Labor Statistics and Petrosky-Nadeau and Zhang (2021).

Note: The figure plots the relationship over time of the between the monthly unemployment rate (from the Current Population Survey) and monthly job openings rate (from Petrosky-Nadeau and Zhang 2021) from January 1959 to December 2023, using different icons and colors to highlight the location of the Beveridge curve at different time periods.

The common textbook interpretation of the Beveridge curve is that it represents where the unemployment rate stabilizes at a given level of the job openings rate, holding everything else constant (for explanations of this interpretation, see Pissarides 2000, chap. 1; Elsby, Michaels, and Ratner 2015). This interpretation, however, runs into serious limitations in practice. As Figure 1 shows, even though the job-openings and unemployment rates tend to move in opposite directions, the negative relationship between them is not stable over time. Thus, at any point in time, it is hard to assess how changes in the number of job openings translate into changes in unemployment.

To understand how unemployment relates to the number of job openings, we need to understand why the Beveridge curve might shift over time. In this paper, we discuss how the relative importance of factors that contribute to movements of the US Beveridge curve has changed over time. We review these factors in the context of

---

of Petrosky-Nadeau and Zhang (2021) goes even further back to 1919, where they use data from the MetLife help-wanted advertising index (available via the NBER macrohistory files) for the first half of the twentieth century.

a flow analogy used to capture the main insights of search and matching theories of the labor market. We then argue that whether the Beveridge curve shifted and why are important considerations if one wants to use the Beveridge curve to understand the potential tradeoffs between inflation and unemployment, as some researchers have done during the COVID pandemic. In particular, how a given scenario affects inflation, job openings, and inflation will depend, among other things, on where along a given Beveridge curve the economy is located and whether the scenario causes the Beveridge curve to shift or not.

To preview the shifting reasons for Beveridge curve shifts, consider Figure 1. Between 1971 and 2000, the curve shifted out and then in, moves that we argue are related to demographic changes in the labor force during that time. Since 2000, the Beveridge curve experienced two significant shifts. The first is a large and persistent outward movement in the curve after the Great Recession of 2007–2009. We link this shift to a long-lasting decline in unemployment outflows that can be attributed to a rise in mismatch between the needs of employers and the skills of the unemployed, a decline in the intensity of recruitment efforts by employers seeking workers, and a rise in the share of long-term unemployment that made it more difficult for workers to leave unemployment. The second, and by far largest, happened after the onset of the COVID pandemic in March 2020, as reflected by circles in Figure 1. We argue that this shift reflects a myriad of factors: a surge in unemployment inflows at the start of the pandemic followed by a rapid recovery and rehiring, as well as a rise in quits and workers switching their employers.

## **Labor Market Frictions and the Beveridge Curve**

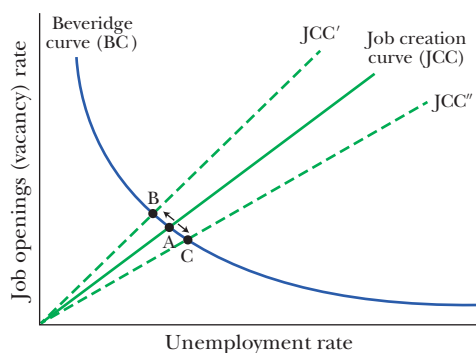
At the heart of the Beveridge curve is the coexistence of unemployment and job openings. That is, at any given moment, there are both workers unable to find jobs and employers looking for suitable workers. The joint existence of these two phenomena implies that “frictions” that prevent workers and employers from being matched instantaneously are crucial to understanding the labor market. Many macroeconomic models of the labor market focus on modeling such frictions. Indeed, the 2010 Nobel prize in economics was awarded to Peter Diamond, Dale Mortensen, and Christopher Pissarides “for their analysis of markets with search frictions” (Royal Swedish Academy of Sciences 2010), and Pissarides (2000) provides a useful overview of this approach as applied to labor markets.

To understand the main intuition for why such models can generate a negative relationship between job openings and unemployment, one can consider the commonly used bathtub analogy for the flow dynamics of the labor market. Let the level of water in a bathtub represent the stock of unemployed. Upward pressures on the number of unemployed are due to people flowing into unemployment, either because they lose their job or quit and look for another one, or because they decide to join the labor force and start looking for work. In the bathtub analogy, one can imagine these flows as the water flow coming from the faucet.

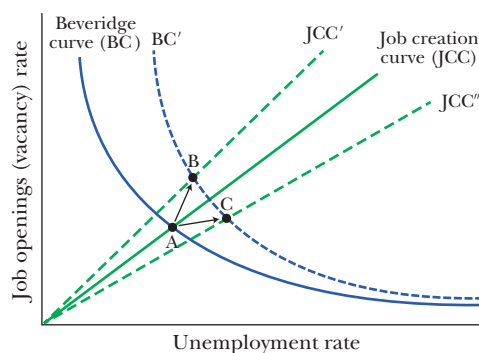
Figure 2

**Theoretical Movements along and of the Beveridge Curve**

Panel A. Beveridge and job creation curves, and equilibrium



Panel B. Joint shifts in Beveridge and job creation curves



Source: Authors' calculations.

Note: The figure shows how the Beveridge curve and job creation curves move and shift in response to changes in the labor market. Panel A shows how changes in the incentives to create new jobs lead to movements of the job creation curve along a fixed Beveridge curve. Panel B shows the combined effect of an outward shift in the Beveridge curve and upward or downward movements in the job creation curve.

Conversely, the water level in the tub is reduced by flows out of it, which is governed by how high the stopper in the drain is raised. The gap created by the stopper can be interpreted as the demand for workers, as measured by the number of job openings. For a given level of job openings, the level of outflows will be higher when the unemployment level is higher. In the bathtub analogy, a higher water level causes more flow pressure on the drain. In models with search frictions, it is easier for employers to find suitable unemployed workers to fill a job opening when there are more unemployed workers.<sup>2</sup>

The water in the bathtub—the stock of unemployed workers—will stabilize when the outflows out of the tub equal the inflows into it. For a given level of inflows, the higher the level of the job openings rate (corresponding to a higher level for the stopper), the lower the level at which the unemployment rate will stabilize.

The Beveridge curve visualizes this level of the unemployment rate as a function of the job openings rate. The solid blue line in panel A of Figure 2 plots such a stylized conceptual Beveridge curve. The downward-sloping curve in panel A of Figure 2 is depicted as steep at low levels of unemployment and flat at high levels. When few workers are unemployed and labor demand is high, it is difficult for firms to hire additional workers, even if they post more job openings. Conversely, when

<sup>2</sup>In theoretical models, the relationship between the number of matches between employers and workers, the unemployment rate, and the job openings rate is often captured by a matching function (Petrongolo and Pissarides 2001).



unemployment is high and there are few job openings, each job opening is likely to be filled quickly, and so reductions in job openings have a substantial impact on unemployment. As a result of these dynamics, a change in the job openings rate has a much smaller effect at a low unemployment rate than at a high one.

The location of the economy along the Beveridge curve is determined by a second relationship, known as the “job creation curve,” which reflects the strength of labor demand (as measured by the job openings rate) at different levels of the unemployment rate. The job creation curve is generally assumed to be upward-sloping for two reasons. First, at a higher unemployment rate, employers tend to be able to hire workers at lower wages. Second, employers find it easier to hire suitable candidates when the unemployment rate is higher simply because there are more people applying per job opening. The solid upward-sloping line in panel A of Figure 2, labeled JCC, plots the job creation curve. We plot the JCC as going through the origin, which is the case in most standard theoretical search models (for example, Mortensen and Pissarides 1994), but that choice does not matter for the rest of our exposition. What matters is that the job creation curve is upward sloping. Most conventional models of search feature an upward-sloping job creation curve, although Yashiv (2016) shows it may be downward sloping in the presence of adjustment costs.

The equilibrium combination of the unemployment and job openings rates that we observe in the data is the one where the job creation curve JCC and the Beveridge curve BC intersect. This corresponds to point A in panel A of Figure 2. Changes in labor demand rotate the job creation curve. If the Beveridge curve was stable and changes in unemployment were solely driven by movements in labor demand, then the observed data from movements in the JCC would trace out this stable Beveridge curve. As shown in panel A of Figure 2, if labor demand increases, the job creation curve shifts counterclockwise and up from JCC to JCC' and the new equilibrium is point B, higher up along the Beveridge curve. Similarly, if labor demand declines, the job creation curve shifts clockwise and down from JCC to JCC'', with a new equilibrium point at point C.

However, as Figure 1 illustrated, US unemployment and job openings rates trace out an empirical Beveridge curve that is far from stable. This was especially true during and after the COVID pandemic. The bathtub analogy provides a useful insight here on why the curve might shift. There are only two things that can change the level in the bathtub conditional on a given job openings rate: (1) a change in inflows into the bathtub; and (2) an (un)clogging of the drain that alters the outflows from unemployment for a given height of the stopper. Translating this to the labor market, factors that alter either the unemployment inflow rate or outflow rate, independent of a change in the job openings rate, will generate shifts in the Beveridge curve.

What is important for understanding the US data plotted in Figure 1 is that joint movements in both the Beveridge and job creation curves can result in the job openings rate and the unemployment rate moving in the same direction, rather than opposite to each other. Consider the example in panel B of Figure 2 where we

evaluate upward and downward shifts in the job creation curve but now explore the case where the Beveridge curve shifts outward from BC to BC'. First, consider the case where the job creation curve moves upward from JCC to JCC'. In this case, the equilibrium moves from point A to point B in the diagram, and the unemployment rate increases a little in spite of a large increase in job openings. Now, consider the case when the Beveridge curve shifts outward from BC to BC' and the job creation curve moves down from JCC to JCC'. This shifts the equilibrium from point A to point C, resulting in a large increase in the unemployment rate even though the job openings rate increases slightly.

The reasons for why the Beveridge curve shifts, and how and whether they also affect job creation, cannot be gleaned from movements in the unemployment and job openings rates alone. Additional evidence needs to be brought to bear. The bathtub analogy suggests that data on labor market flows may be able to offer key insights on the reasons for why the Beveridge curve might have shifted.

## Reasons for Beveridge Curve Shifts over Time

We focus on three recent episodes for which we have detailed labor market data that allow us to illustrate the main drivers of the changing position of the US Beveridge curve in the context of the inflow-outflow framework from our bathtub analogy. Figure 1 highlights the three episodes on which we focus: a back-and-forth shift of the Beveridge curve between 1970 and 2009; a rightward shift in the wake of the Great Recession; and the shifts during and after the COVID recession. To accentuate medium-term movements in the Beveridge curve, Figure 1 plots each episode with a different marker.

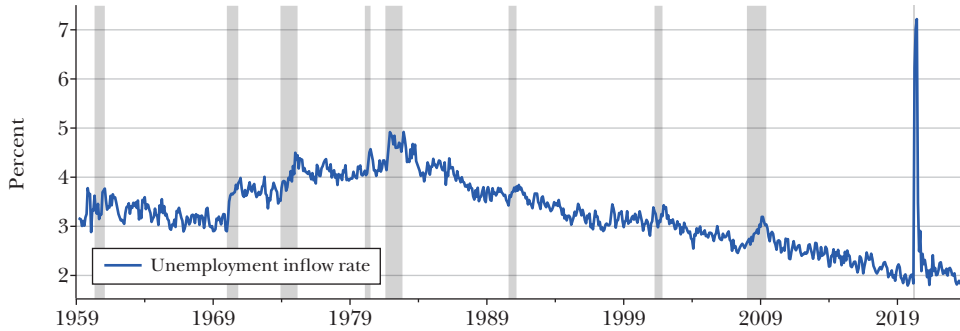
### The “Grand Gender Convergence” and Aging of the Baby Boomers: 1970–2009

Starting in 1970, the US Beveridge curve moved rightward by about 3 percentage points along the horizontal unemployment rate axis and stayed there throughout most of the 1980s, as shown in Figure 1. Then, in the late 1980s, the Beveridge curve moved back to the left, ending up further left in the 2000s than it started 30 years before. These medium-term movements in the Beveridge curve were driven by unemployment inflows. Figure 3 shows the time series of an estimate of the monthly inflow rate into unemployment and outflow rate from unemployment, constructed using the methodology described in Shimer (2012).

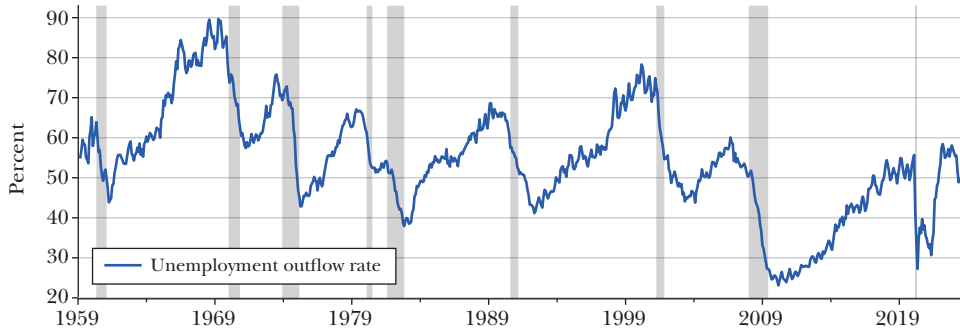
Two patterns in these series are worth noting for their effect on the Beveridge curve. First, unemployment inflows tend to spike during the initial stages of recessions, which puts upward pressure on the unemployment rate. However, these spikes tend to have only a short-lived effect on the position of the Beveridge curve. Secondly, there have been sizable medium-run movements in the rate of inflow into unemployment. It trended up from the late 1960s through the mid-1980s. After that, the inflow rate into unemployment trended down, and its level in 2023 lies below its level in 1960.

Figure 3  
**Unemployment In- and Outflow Rates**

Panel A. Unemployment inflow rate



Panel B. Unemployment outflow rate



Source: US Bureau of Labor Statistics and authors' calculations.

Note: Three-month moving average of the unemployment inflow rate (as a percent of the employed) and outflow rate (as a percent of the unemployed) estimated using the methodology from Shimer (2012).

In turn, the up and down trends in unemployment inflows can largely be traced to two important and closely intertwined factors: (1) the “Grand Gender Convergence”; and (2) the entry of the “baby boom” generation into the US labor force and its subsequent aging.

During the Grand Gender Convergence, the US economy experienced an increase in female labor force participation from around 47 percent in 1976 to 60 percent in 2000 (for a detailed discussion of the changing employment status of US women over time, see Goldin 2006). This influx of labor market participants, as they entered the labor force and later on as they reentered after leaving the labor force temporarily due to family responsibilities, initially put upward pressures on the unemployment inflow rate. However, as social norms changed and the availability of maternity leave increased in the late 1970s and 1980s, employment relationships of women became more stable. Women experienced fewer career interruptions, lowering the rate at which they dropped out of the labor force and subsequently flowed back in as unemployed, in this way reducing the unemployment inflow rate.

The rise and subsequent decline in the unemployment inflow rate for women coincided with the entry and aging of the “baby boom” cohort born between 1946 and 1963. In the 1970s, most of this generation consisted of teens and people in their twenties and, as usual with younger workers, had less-stable jobs and more frequent spells of unemployment. This contributed to higher unemployment inflows. As baby boomers entered their prime working ages and settled in to more steady jobs in the 1980s and 1990s, their inflow rate into unemployment trended down.

To see why the medium-term movements in unemployment inflows shifted the empirical Beveridge curve as shown in Figure 1, reconsider panel B of Figure 2. In that example, an increase in unemployment inflows is captured by a rightward shift of the Beveridge curve, from BC to BC'. But in addition, higher unemployment inflows due to less stable employment relationships reduce the expected duration of job matches and, with that, the incentive to create jobs. In the context of the example in the figure, the increase in the unemployment inflow rate between 1970 and 1985 can be interpreted as a clockwise rotation of the job creation curve from JCC to JCC'.<sup>3</sup> The result is that the equilibrium moves from point A to point C in panel B of Figure 2. This reflects an unambiguous increase in the unemployment rate, while the sign of the change in job openings is undetermined. Several studies, including Shimer (1998), discuss the importance of demographic trends for the rise in the unemployment rate in the 1970s and 1980s and its decline in the 1990s.

A downward trend in the unemployment inflow rate started in the mid-1980s, and by the 2000s, the unemployment inflow rate was lower than in the 1970s, as documented in Crump et al. (2019). Consistent with this pattern, the empirical Beveridge curve in the 2000s was located to the left of the one in the 1970s (as one can see in Figure 1). However, even though unemployment inflows continued their downward trend into the 2010s, the Beveridge curve did not shift further inwards. Instead, it moved outwards in the wake of the Great Recession that started in 2007.

### **The Great Recession and Match Efficiency: 2010–2019**

The persistent outward shift of the empirical Beveridge curve in the 2010s, as shown in Figure 1, cannot be explained by changes in inflows to unemployment, which were only elevated for the better part of two years during the recession (as shown in panel A of Figure 3). The persistent shift must instead be traced back to a decline in outflows from unemployment.

In the wake of the Great Recession, the labor market seems to have become less efficient in matching unemployed workers with available job openings. To see this, consider panel B of Figure 3, which shows the time series of the unemployment outflow rate. Prior to 2009, the outflow rate from unemployment followed a regular procyclical pattern. These fluctuations coincided with cyclical changes in the job

<sup>3</sup>Elsby, Michaels, and Ratner (2015) provide a mathematical example of how an increase in the unemployment inflow rate both shifts the Beveridge as well as the job creation curve. They also provide a counterfactual empirical Beveridge curve that takes out the impact of variations in unemployment inflows.

openings rate (not shown in the figure) and can be interpreted as movements in job creation along a relatively stable Beveridge curve, as illustrated in panel A of Figure 2. By contrast, the decline in the unemployment outflow rate during and after the Great Recession was unusually large, both by historical standards and as compared to the drop in job openings (as shown in Elsby, Hobijn, and Şahin 2010).

Before we dive into the possible causes of this atypical decline in unemployment outflows, it is useful first to consider what our framework from Figure 2 teaches about how such a decline should affect the observed Beveridge curve (the discussion here draws from Daly et al. 2012). Using the bathtub analogy, an unusual decline in unemployment outflows for a given job openings rate can be interpreted as a clogging of the drain. In the labor market, this corresponds to a reduction in the productivity of the matching process, often referred to as a decrease in matching efficiency. During and after the Great Recession, this type of a decline pushed the Beveridge curve rightward, from BC to BC' in panel B of Figure 2. Moreover, as it became harder for employers to fill job openings, it also became less attractive to post job openings. This is reflected in a clockwise and downward rotation of the job creation curve from JCC to JCC'. Just like in the case of the increase in the inflow rate we covered above, the net effect is that the equilibrium moves from point A to point C in panel B of Figure 2. This again results in an unambiguous increase in the unemployment rate.

Economists have sought to identify the source of the persistent decline in matching efficiency and have offered several reasons behind it (for example, see Daly et al. 2012; Elsby, Michaels, and Ratner 2015, and references therein). We discuss four key potential contributors: (1) repeated extensions of unemployment insurance benefits may have affected the job search behavior of the unemployed; (2) firms may have increased their hiring standards, sometimes described as a decline in “recruiting intensity”; (3) mismatch between the locations and qualifications of the unemployed relative to available job openings may have increased; and (4) the unemployed may have experienced the “scarring” effect of the recession.

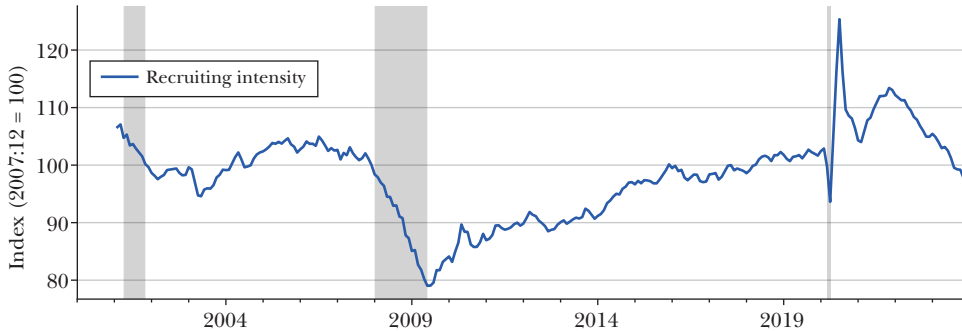
The first channel posits that the unemployed put less effort into finding a job when they can rely on extended unemployment benefits. While this sounds plausible, several papers found this effect to be quantitatively unimportant (for example, see Farber, Rothstein, and Valletta 2015; Farber and Valletta 2015; Chodorow-Reich, Coglianesi, and Karabarbounis 2019). In fact, search effort by the unemployed reached an all-time high during and after the Great Recession as documented by Mukoyama, Patterson, and Şahin (2018), which acted to moderate the outward shift in the Beveridge curve instead of accounting for it.<sup>4</sup>

The second potential channel is the decline in firms' recruiting effort, which one could interpret as the flipside of workers' search effort. In our conceptual framework, job openings capture firms' hiring efforts. However, firms may use

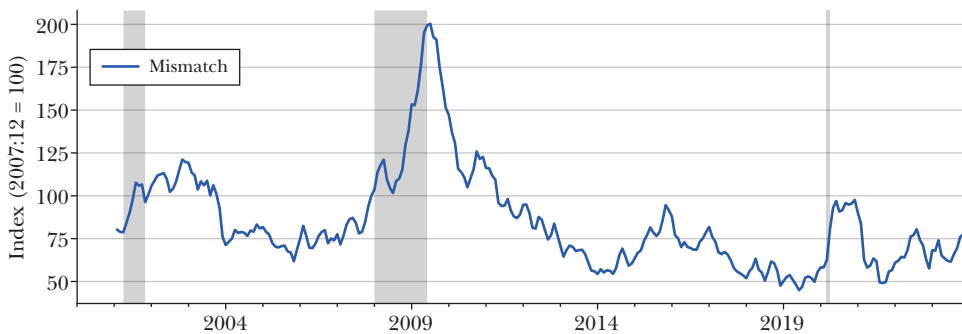
<sup>4</sup>Mukoyama, Patterson, and Şahin (2018) generate an index of search effort from reported job search methods in the Current Population Survey (CPS) and time spent on job search reported in the American Time Use Survey. They show that this index is countercyclical, with a sharp rise during the Great Recession period.

Figure 4  
Mismatch and Recruiting Intensity

Panel A. Recruiting intensity



Panel B. Mismatch



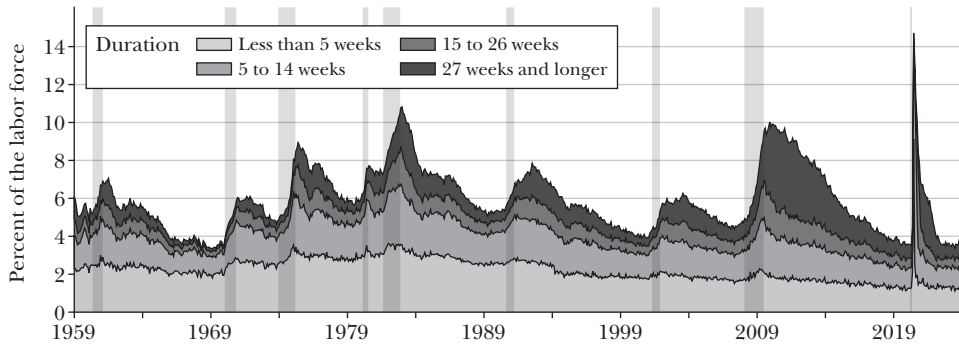
Source: Authors' calculations.

Note: Three-month moving averages of monthly seasonally-adjusted time series, measured as an index Dec-2007=100. Recruiting intensity constructed using method from Davis, Faberman, and Haltiwanger (2013). Mismatch index calculated as described in Şahin et al. (2014).

other margins besides the number of job openings posted to adjust their hiring. For example, they can allocate fewer resources to recruiting, increase their hiring standards, or cut back on benefits. Several studies stress the importance of these additional margins in firms' recruiting processes and construct a measure of recruiting effort for firms, which is often labeled as "recruiting intensity." We use an updated index of recruiting intensity developed by Davis, Faberman, and Haltiwanger (2013). They derive their measure from the Job Openings and Labor Turnover Survey (JOLTS) hiring rate, which they interact with their estimated elasticity of employers' job-filling rates with respect to hires.<sup>5</sup> Panel A of Figure 4 shows that the outward shift in the Beveridge curve coincided with a decline in recruiting intensity after the Great Recession. Formal quantitative analyses suggest that the reduction

<sup>5</sup> See Hershbein and Kahn (2018), Modestino, Shoag, and Ballance (2020), and Carrillo-Tudela et al. (2023) for alternative measures of recruiting intensity, such as hiring standards listed in online job ads and hiring standards, search effort, and wage offers reported in firm recruitment surveys.

Figure 5

**Duration Distribution of the Unemployed**

Source: US Bureau of Labor Statistics and authors' calculations.

Note: The figure plots the percent of the labor force that has been unemployed by different durations, as calculated from the Current Population Survey. The shares are stacked to cumulate to the total unemployment rate.

in recruiting intensity accounts for about 2 percentage points of the unemployment rate in 2010 and 1 percentage point of the unemployment rate in 2012 for the post–Great Recession shift in the Beveridge curve (Crump et al. 2019).

A third channel to consider is “mismatch,” which refers to the misalignment between the skills or locations of unemployed workers and available job openings. An increase in mismatch would lower the outflow rate, because matching unemployed workers to job openings would be more difficult. Considering that around half of the job losses during the Great Recession were concentrated in just two industries—construction and manufacturing—an increase in skill mismatch at this time seems likely. Moreover, house prices declined drastically, which made it harder for unemployed workers to move to pursue job opportunities. Panel B of Figure 4 shows a measure of skill mismatch from Şahin et al. (2014). Their mismatch index provides a summary measure of differences between each industry’s unemployment and job openings shares. We update their index using industry-level unemployment data from the Current Population survey and job openings data from the Job Openings and Labor Turnover Survey. The sharp rise in skill mismatch is evident during the Great Recession. Şahin et al. (2014) find that this increase contributed about 1 percentage point to the horizontal shift of the Beveridge curve in terms of the unemployment rate in 2010. Interestingly, they find a quantitatively minor role for geographic mismatch.

The final potential channel for the decline in outflows from unemployment is that the Great Recession was a major disruption for careers, with a record fraction of the labor force remaining out of a job for more than six months, as seen in Figure 5, which shows the duration composition of the unemployed. Long-term unemployment can result in a loss of job-related skills and networks and a negative stigma when searching for a new job. These “scarring” effects diminished the

reemployment prospects of a large fraction of the unemployed. As a result, the outflow rate remained depressed even as the economy recovered. A more detailed discussion of the “scarring” effect on the Beveridge curve can be found in Elsby, Michaels, and Ratner (2015).

### **Pandemic-Related Factors: 2020–2023**

The movements in the empirical Beveridge curve after the Great Recession pale in comparison to those during and after the outbreak of the COVID pandemic from 2020 to 2023, as shown in Figure 1. While previous shifts in the Beveridge curve were clearly attributable to changes in either unemployment inflows or outflows, the time period during and after the pandemic involved both. Here, we discuss four different phases of pandemic-related factors that affected the empirical Beveridge curve, illustrated in Figure 6.

The first phase is the onset of the pandemic, which coincided with broad-based lockdowns. When COVID broke out in March 2020, employers laid off a record number of workers. Nonfarm payrolls declined by more than 20 million jobs. Because it was not clear how long the pandemic would last, many of these layoffs were temporary, with an explicit intention to rehire workers as circumstances improved. The record number of layoffs resulted in an unprecedented spike in unemployment inflows that can be seen in the top panel of Figure 4. This resulted in a large outward shift in the Beveridge curve and a drastic decline in job creation, reflected by the clockwise rotation of the job creation curve illustrated in panel A of Figure 6. As a consequence, the unemployment rate rose to a post-World War record of 14.7 percent. The job openings rate barely dropped.

At first glance, it might seem surprising that the job openings rate did not drop much more given the drastic reduction in labor demand during the pandemic. However, panel A of Figure 6 provides a clear intuition for why this was the case. The change in the job openings rate is ambiguous when the Beveridge curve shifts outward and job creation declines. This reveals an important insight: changes in the job openings rate only reflect changes in labor demand if the Beveridge curve is stable.

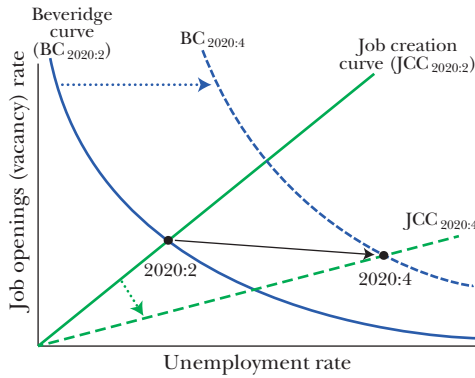
This drastic shift at the beginning of the pandemic was short-lived. During the summer and fall of 2020, many of the workers that were laid off during the lockdown were recalled as the economy gradually reopened. Consequently, the unemployment rate retreated to 6.3 percent at the beginning of 2021 without a notable increase in the job openings rate; Hall and Kudlyak (2022) discuss the importance of temporary layoffs during the COVID pandemic. The latter half of 2020 represents a unique time period with large unemployment outflows driven by high matching efficiency as workers were “rematched” with their former employers. Panel B of Figure 6 depicts this period. It corresponds to a partial reversal of the rightward shift in the Beveridge curve. This reversal was only partial because the change in composition of demand away from in-person services created a gap between the needs of employers and skills of those unemployed. Consistent with this, mismatch was elevated in 2020 and early 2021, as can be seen in panel B of Figure 4. A limited renormalization of daily life resulted in some recovery in economic activity and,



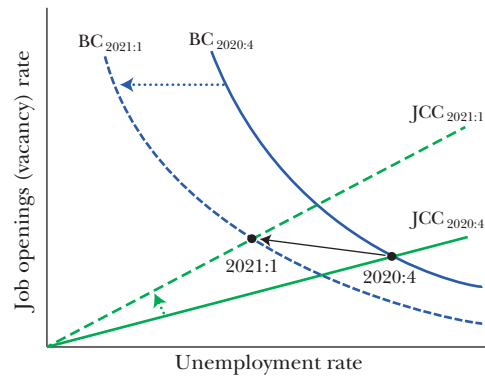
Figure 6

The Covid Episode and the Great Resignation in Beveridge Space

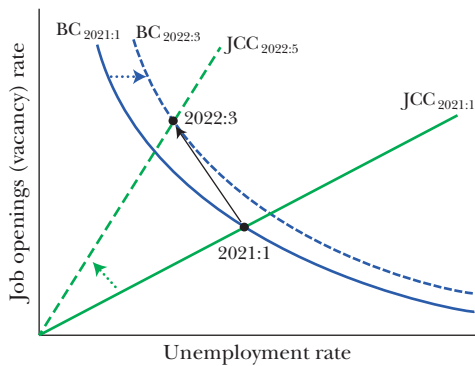
Panel A. 2020:2–2020:4: Spike in layoffs from lockdowns



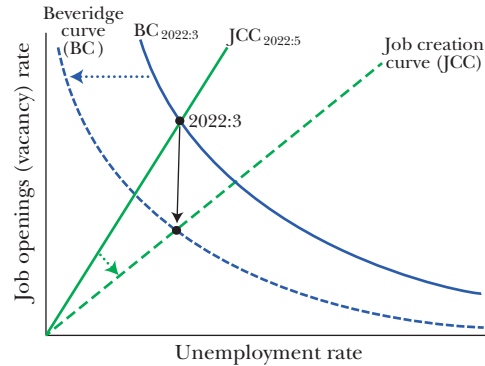
Panel B. 2020:4–2021:1: Recall hiring during reopening



Panel C. 2021:1–2022:3: Great Resignation and job-to-job transitions



Panel D. 2022:3–2023:12: Cooling of the labor market



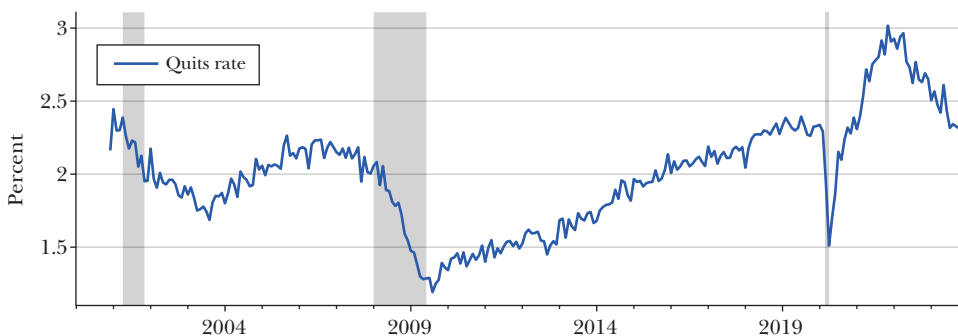
Source: Authors' calculations.

Note: The figure plots shifts in the Beveridge curve and job creation curve during different the effect of phases of the Covid pandemic and recovery. Panel A captures the outward shift of the Beveridge curve and decline in job creation at the start of the pandemic. Panel B captures the partial undoing of the initial shock as the economy reopened. Panel C captures the shift out in the Beveridge curve and increase in job creation during the Great Resignation. Panel D captures the reversal of the Great Resignation.

with it, the demand for workers. This pattern is captured by a counterclockwise upward rotation of the job creation curve in our framework.

The shifts in the Beveridge curve starting in 2021 are more puzzling than those in 2020. The job openings rate rapidly climbed from 4.8 percent in December 2020 to a high of 7.4 percent in March 2022, while the unemployment rate declined from 6.3 to 3.6 percent—one of the lowest readings in half a century. This period was followed by a decline in the job openings rate without a notable increase in unemployment, which

Figure 7  
**Quits Rate**



Source: US Bureau of Labor Statistics and authors' calculations.

Note: The figure plots the three-month moving average of the monthly quit rate (as a percent of employment) for nonfarm employment, as calculated from the Job Openings and Labor Turnover Survey.

continued throughout 2022 and 2023. Our framework provides a potential narrative of these unusual movements in the Beveridge curve by recognizing the importance of another pandemic-related labor market development: the “Great Resignation.”

Starting in 2021, the US labor market saw an unprecedented rise in quits, as many workers, after reevaluating their career choices and work-life balance, decided to switch jobs. Figure 7 shows the number of quits in a month as a percentage of total employment, known as the quits rate. It reached an all-time high of 3.0 percent in the spring of 2022, well above its pre-COVID average of 1.9 percent in the Job Openings and Labor Turnover Survey data. When more employed workers are actively looking for new job opportunities, it affects the job-finding prospects of the unemployed workers as well as job-posting incentives of firms. First, for a given level of job openings, the job-finding rate of the unemployed falls because they compete with the employed for jobs. The result is an outward shift in the Beveridge curve. At the same time, firms' decisions to post job openings are positively affected when there are more employed looking to change jobs. Therefore, the job creation curve rotates further counterclockwise. Panel C of Figure 6 illustrates the effects of the on-the-job search channel in Beveridge space between early 2021 and spring 2022. It provides intuition behind why the recovery in labor demand, together with the Great Resignation, resulted in a stark increase in the job openings rate and a decline in the unemployment rate to 3.6 percent by March 2022. It is also consistent with Elsy, Michaels, and Ratner (2015), who provide a formal theoretical treatment of the impact of on-the-job search on the Beveridge and job creation curves.

Starting in the summer of 2022, there was a puzzling drop in job openings without any change in the unemployment rate. However, if labor market developments since early 2022 are interpreted as a partial reversal of the shifts in the Beveridge and job creation curves since the start of 2021, then the drop no longer

appears perplexing. Panel D of Figure 6 illustrates the intuition. Starting in early 2022, the quits rate declined, which we take as an indication of a reduction in workers' eagerness to switch jobs. This contributed to a clockwise, downward rotation of the job creation curve and an inward shift of the Beveridge curve. If the horizontal movements in the two curves offset each other, they combine to generate a vertical drop in job openings while the unemployment rate remains constant.

The recovery from the pandemic was different from the previous two episodes we considered. It was characterized not only by a rapid recovery in labor demand but also by a surge in workers actively looking for new jobs. The latter likely reduced the efficiency with which the unemployed were matched with job openings and may have induced employers to post more of them. The combined effects can only be understood if one takes into account the implied joint movements of the Beveridge and job creation curves.

## **Beveridge Curve Shifts and the Unemployment-Inflation Tradeoff**

The COVID pandemic generated a renewed interest in the Beveridge curve, with several papers arguing the Beveridge curve framework could provide useful information about the tradeoff between unemployment and inflation (Blanchard, Domash, and Summers 2022; Figura and Waller 2022). As central banks raised nominal interest rates to rein in high inflation during the post-COVID economic recovery, a key question was whether bringing down inflation necessitated a large rise in unemployment or whether a “soft landing,” with little or no rise in unemployment, was possible. Studies that relied on the Beveridge curve to address the question reached different conclusions on the answer.

After these studies came out in mid-2022, inflation began to fall without much change in unemployment. Part of the fall in inflation appeared to be driven by factors specific to the recovery from the pandemic that were unrelated to the labor market, such as improvements in global supply chain disruptions. Nevertheless, one can still ask how the Beveridge curve is related to inflation more generally, independently of why inflation fell in the post-pandemic period. In this section, we argue that if the Beveridge curve is stable, whether raising the nominal interest rate can achieve a soft landing depends on the shape of the Beveridge curve. If the Beveridge curve shifts, whether inflation will fall for a fixed unemployment rate depends on the reason(s) the Beveridge curve shifted.

### **A Framework for Analyzing Inflation**

The starting point for analyzing inflation is not the Beveridge curve but another curve named after the economist who conceived it—the Phillips curve. While Phillips (1958) studied the relationship between unemployment and nominal wage growth, the curve bearing his name has since come to refer to any relationship between some measure of economic activity, like unemployment or output, and some measure of price changes, like wage growth or inflation. A Phillips curve

relationship between output and inflation is one of the key pillars of the textbook New Keynesian model commonly used to study monetary policy.<sup>6</sup>

To understand this textbook Phillips curve, consider what happens when a central bank sets a temporarily higher nominal interest rate. The effect of this policy on output and inflation depends on the ability of producers to change their prices in response to the higher nominal cost of borrowing. If firms can adjust prices in full, inflation would rise in tandem with the nominal interest rate. The real interest rate (the nominal interest rate net of inflation) would be unchanged. In turn, full price adjustment should leave aggregate demand unaffected: demand for goods and services in principle depends on the real interest rate, or the amount of consumption we need to give up tomorrow to consume today. Thus, in a frictionless world, increasing the nominal interest rate would raise inflation but would have no effect on the real interest rate, aggregate demand, or output.

In practice, firms may not adjust prices immediately or fully in response to a higher nominal interest rate. In this case, inflation would no longer rise in tandem with the nominal interest rate. The real interest rate would have to rise, dampening aggregate demand and reducing output. If firms produce fewer goods, their marginal cost of production would fall; for example, employers might be able to offer lower wages if they need to attract fewer workers to produce. The lower marginal costs of production would lead some firms to set lower prices. In the short run, a higher nominal interest rate would lead to a higher real interest rate, lower aggregate demand and output, lower labor demand, and lower inflation. Eventually, firms would fully adjust their prices, and the real interest rate, aggregate demand, and output would return to their long-run levels—that is, the levels that prevail when prices are fully flexible.

Conceptually, the Phillips curve captures how aggregate demand affects inflation. However, one cannot measure aggregate demand directly. In practice, economists have used variables that tend to co-move with aggregate demand, such as the unemployment rate, to study the Phillips curve empirically.

### **Relating Unemployment and Inflation with a Stable Beveridge Curve**

The textbook model based on a Phillips curve between output and inflation abstracts from unemployment. To relate aggregate demand to unemployment, we can turn to the Beveridge curve. When firms cannot fully adjust their prices, a higher nominal interest rate would lower aggregate demand, leading firms to require less labor. Demand for labor determines the location of the job creation curve. A higher nominal interest rate would shift the job creation curve down. If the Beveridge curve remained stable, such a move would lead to higher unemployment and fewer job openings. For a formal analysis of how to incorporate unemployment into the textbook New Keynesian model, see Blanchard and Galí (2010) and Galí (2011).

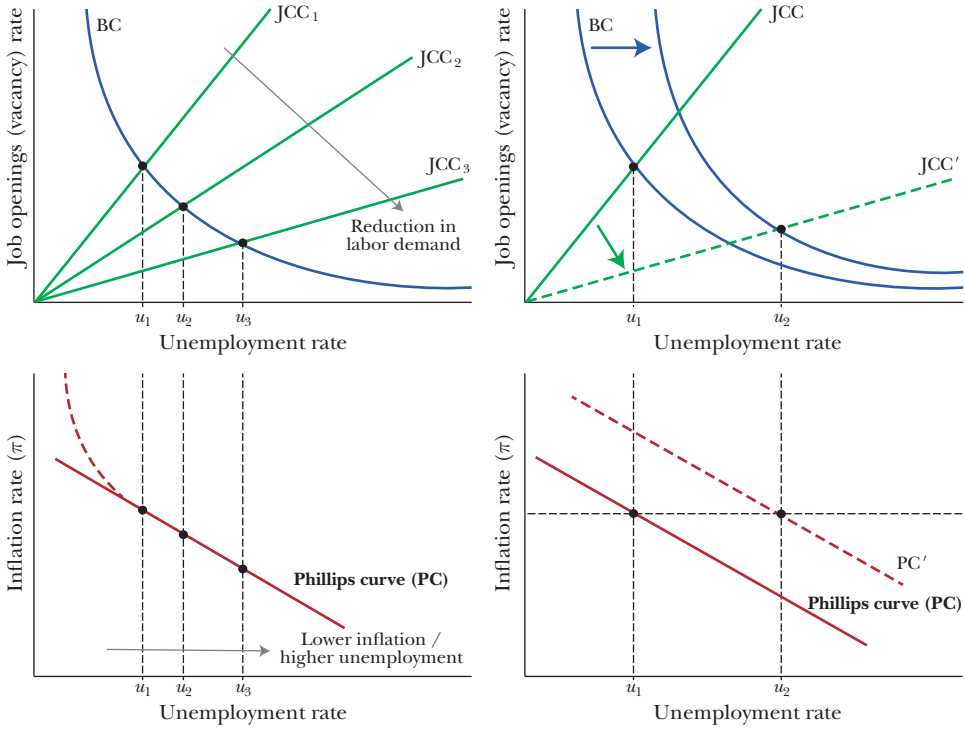
<sup>6</sup>More precisely, the Phillips curve in the New Keynesian model involves the output gap, or the ratio of output to the longrun level of output that would prevail when prices are fully flexible. For a discussion, see Woodford (2003) and Galí (2015).

Figure 8

The Link between the Beveridge and Phillips Curves

Panel A. Beveridge, job creation, and Phillips curves

Panel B. Increase in long-run unemployment



Source: Authors' calculations.

Note: The figure shows the connection between the Beveridge curve and Phillips curve. Panel A shows how tighter monetary policy causes a shifts down in the job creation curve down along a fixed Beveridge curve raises unemployment as it reduces inflation. Panel B shows how an outward shift in the Beveridge curve and a fall in job creation would be associated with a shift to right in the Phillips curve.

We can use the Beveridge curve to determine the unemployment rate for different nominal interest rates, as illustrated in Figure 8. Panel A shows how a stable Beveridge curve at the top links to a stable Phillips curve at the bottom. As the central bank increases the short-term nominal rate, it will both lower inflation and move the job creation curve down along the Beveridge curve. This will trace out a downward-sloping Phillips curve relationship between inflation and the unemployment rate.

Researchers have estimated versions of the curve depicted in the bottom of panel A of Figure 8 with pre-COVID data and found it is downward-sloping but

relatively flat with respect to unemployment.<sup>7</sup> This finding suggests that if a central bank temporarily increased the nominal interest rate to achieve a significant decline in inflation, it would have to tolerate a substantial increase in unemployment to bring down inflation, denying the possibility of a soft landing.

However, estimates of a flat Phillips curve are based on pre-COVID data. During the recovery from COVID, the unemployment rate fell below its pre-pandemic level. Recall that the Beveridge curve should in principle be steep at low levels of unemployment: when few workers are unemployed, posting more job openings cannot translate into a large change in unemployment. A shift in the job creation curve would then primarily affect job openings rather than unemployment. The implication would be that at low levels of unemployment, the Phillips curve would also be steeper, as illustrated by the dashed line in the bottom of panel A of Figure 8, rather than flat as depicted by the solid line. Moving the job creation curve along the fixed Beveridge curve would then primarily affect job openings rather than unemployment, and it should be possible to lower inflation without a large rise in unemployment after all. Figura and Waller (2022) discuss and estimate the shape of the Beveridge curve at low unemployment rates.

In short, if the Beveridge curve remains stable, whether a central bank can bring down inflation without raising unemployment depends on the steepness of the Beveridge and Phillips curves. But as we discussed in the previous section, the Beveridge curve has not remained stable over time. This raises the question of how shifts in the Beveridge curve matter for the tradeoff between inflation and unemployment.

### **Implications of Shifts in the Beveridge Curve for the Phillips Curve**

Some economists have argued that to account for the Beveridge curve in studying inflation, we can replace the unemployment rate in the Phillips curve with the ratio of job openings to unemployment (for example, Ball, Leigh, and Mishra 2022; Bernanke and Blanchard 2023; and Crust, Lansing, and Petrosky-Nadeau 2023). This approach incorporates both variables that constitute the Beveridge curve. The ratio between the two is the key equilibrium object in frictional search models, and using this ratio allows both variables to matter for inflation. However, simply replacing unemployment with the ratio of job openings to unemployment is either unnecessary or only partly incorporates how the Beveridge curve matters for inflation.

If the Beveridge curve remains stable, using the ratio of job openings to unemployment rather than the unemployment rate should not matter. For a stable Beveridge curve, each unemployment rate is associated with a unique ratio of job openings to unemployment, and the two should be equally informative about the

<sup>7</sup>Just as the Beveridge curve shifts, so does the Phillips curve. Estimating a Phillips curve between inflation and unemployment requires accounting for these shifts—for example, by controlling for changes in inflation expectations and in the natural rate of unemployment. See Crump et al. (2019) as one example that accounts for changes in these variables.

state of the labor market. Indeed, Furman and Powell III (2021) confirm that for pre-pandemic data, inflation forecasts using unemployment are indistinguishable from those using the job openings-to-unemployment ratio.

If the Beveridge curve shifts, a Phillips curve that uses the ratio of job openings-to-unemployment will only be useful if the relationship between inflation and the ratio of job openings to unemployment remains fixed despite the shift. We argue this need not necessarily be the case. Consider an outward shift in the Beveridge curve due to a higher separation rate or a decline in match efficiency. In the long run, once firms can adjust their prices in full, the central bank can affect inflation but not the real interest rate, aggregate demand, or output. For any given Beveridge curve, this means that once firms fully adjust their prices in response to monetary policy, aggregate demand would settle down to some long-run level. Suppose the Beveridge curve were to shift out from  $BC$  to  $BC'$  as in the top of panel B of Figure 8. If the separation rate is higher or match efficiency is lower, posting job openings would be less profitable for employers. Employers would post fewer job openings at any given unemployment rate, meaning the job creation curve will move clockwise and fall. The job creation curve will shift from  $JCC$  to  $JCC'$ , as illustrated in panel B of Figure 8. If the changes in the separation rate or match efficiency persist, the outward shift in the Beveridge curve will lead an increase in the long-run unemployment rate from  $u_1$  to  $u_2$ .

If the central bank has a constant long-run target for inflation (as most central banks do in practice), the long run unemployment rates  $u_1$  and  $u_2$  should both be associated with the same target inflation rate. Because the Phillips curve traces short-run deviations in unemployment from its long-run level, a shift in the Beveridge curve that raises the long-run unemployment rate will shift the Phillips curve up to a new downward-sloping curve that reflects the higher long-run unemployment rate.<sup>8</sup> This shift is illustrated at the bottom of panel B in Figure 8. If one were to study a Phillips curve relating inflation to the ratio of job openings-to-unemployment, it would also shift, as the long-run ratio of job openings to unemployment would be lower when the Beveridge curve shifts. A fixed Phillips curve that relates inflation to the ratio of job openings to unemployment would fail to capture the effect of the Beveridge curve shift on inflation.<sup>9</sup>

<sup>8</sup>Crump et al. (2022) provide independent evidence that long-run unemployment early in the pandemic rose in a way that is consistent with a shift up of the Phillips curve as depicted in panel B of Figure 8. They use data on unemployment, inflation, inflation expectations, and wage growth (as a measure of nominal costs) to infer long-run unemployment. They do not use data on job openings or make any assumptions on whether or why the Beveridge curve shifted. Using the same approach, Crump et al. (2019) find that the previous shift of the Beveridge curve during the Great Recession was also associated with higher long-run unemployment due to rising mismatch and declining recruiting intensity. They argue that this, along with anchored inflation expectations, can help explain the “missing disinflation” at the time—that is, why inflation barely fell despite a large rise in unemployment.

<sup>9</sup>A fall in match efficiency will shift the Phillips curve up whether we use unemployment or the ratio of job openings to unemployment. For other shocks that can shift the Beveridge curve, the effect on the Phillips curve depends on whether it is specified using unemployment or the ratio of job openings to unemployment. For example, an increase in the willingness of workers to search on the job will shift the

As we discussed in the previous section, the Beveridge curve shifted out and then back in during and after the pandemic. If these shifts were driven by changes in the separation rate and match efficiency, we should expect similar upward and downward shifts in the Phillips curve. A downward shift would mean lower inflation holding the unemployment rate fixed, allowing for a soft landing. This soft landing would not be due to changes in the nominal interest rate but to the downward shift in the Beveridge curve.<sup>10</sup> However, we argued in the previous section that shifts during the pandemic may have been driven by an increase in the willingness of employed workers to search for new jobs (as illustrated in panels C and D of Figure 6) rather than by changes in the separation rate or match efficiency. As we now discuss, the implications of this scenario for inflation can be different.

### **Implications of an Increase in On-the-Job Search**

Consider an increase in the willingness of workers to search while employed, as illustrated in panel C of Figure 6. One immediate difference from our example in panel B of Figure 8 is that the job creation curve shifts up rather than down; that is, employers find posting job openings more profitable when they can more easily hire workers from other jobs, rather than less profitable when it is harder to match with unemployed workers. The change in long-run unemployment in response to increased on-the-job search is ambiguous: more job openings will make it easier for the unemployed to find jobs, but the unemployed also face more competition from the need to compete with more employed searchers for these jobs. The implication of greater on-the-job search for the Phillips curve will be similarly ambiguous, because changes in the long-run unemployment rate will no longer be the only factors that influence the Phillips curve.

Understanding the implications of on-the-job search for inflation is at the frontier of current macroeconomic research. Recent work has argued that with on-the-job search, the relationship between inflation and economic activity depends on factors beyond unemployment (see Moscarini and Postel-Vinay 2023; Faccini and Melosi 2023). These factors include how many employed workers are searching for new jobs and the intensity with which they search. For reasons we discuss next, changes in the search behavior of the employed can appear as shifts of the Phillips curve.

One aspect of recent models with on-the-job search is the assumption that wages are determined through a process of “offer matching”—that is, firms can match the wage offers that workers receive from other employers. In these models, workers increase their wages in two ways. First, they can find a more productive

---

Beveridge curve out and the job creation curve up. This will increase the long-run ratio of job openings to unemployment but will have an ambiguous effect on long-run unemployment. As a result, the shift in the Phillips curve will be ambiguous if it is specified in terms of unemployment, but not if it is specified in terms of the ratio of job openings to unemployment.

<sup>10</sup>Blanchard, Domash, and Summers (2022) also observed that a downward shift of the Beveridge curve could lead to lower inflation without a rise in unemployment, but they viewed such shifts as unlikely to occur in practice.



employer who is willing to pay them more. Second, they can find outside offers from firms that are less productive than their current employer and force their current employer to pay them a higher wage to retain them. When workers move to more productive employers, both their productivity and their wages rise. When workers get raises from their current employer using outside offers, their wages rise while their productivity remains unchanged. Because inflation depends (in part) on changes in labor costs, the latter scenario will be associated with higher inflation.

When workers are more willing to search for new jobs and employers respond by posting more job openings, we would expect more workers to move to jobs in which they are more productive.<sup>11</sup> As workers move to more productive jobs, they will be less likely to easily find even more productive jobs. Most of their wage growth will then likely come from their current employers in response to outside offers. This means wage growth is more likely to be associated with rising labor costs. Even without any changes in unemployment, we should observe larger increases in labor costs and higher inflation. This would look like an upward shift of the Phillips curve. Hence, an increase in the willingness of workers to search on the job that would appear as an outward shift of the Beveridge curve would also appear as an upward shift of the Phillips curve.

These inflationary pressures can persist even after workers succeed in finding better jobs and employers post fewer job openings. That is, even in the phase corresponding to panel D in Figure 6, where job openings decline and the Beveridge curve shifts inward, inflationary pressures can remain if the workers who moved to more productive jobs continue to receive outside offers. Eventually, these workers will receive enough outside offers to align their wages with their productivity and labor costs will cease to rise, eliminating these inflationary pressures. The key insight is that these additional offers may cause the Phillips curve to shift down with a lag relative to the inward shift of the Beveridge curve, creating a lag between the fall in inflation and the decline in job openings. This is in contrast to a more immediate fall in inflation that would occur had the Beveridge curve shifted inward because of a fall in the separation rate or an improvement in match efficiency. Thus, the timing of the link between changes in inflation and changes in the labor market will depend on why the Beveridge curve shifted.

### **Taking Stock**

The Beveridge curve is a useful tool that researchers and policymakers can use to study unemployment as well as to relate unemployment (and job openings) to inflation under various scenarios. How unemployment and inflation vary depends on the relevant changes in the labor market and how they affect the Beveridge curve. When the Beveridge curve is stable, whether an increase in the nominal interest rate intended to lower inflation will primarily raise unemployment or lower

<sup>11</sup> Barlevy (2002) discusses the notion that greater job posting would lead workers to move towards better matches, although in his paper the higher job openings rate is due to improved productivity rather than a greater willingness by workers to search on the job.

job openings depends on the shape of the Beveridge curve. When the Beveridge curve shifts, whether inflation will fall for a given level of the unemployment rate depends on the reason why the Beveridge curve shifted. As we have discussed, the reasons the Beveridge curve shifted have varied over time. To understand the potential tradeoff between inflation and unemployment in any particular period thus requires additional data beyond unemployment and job openings to ascertain why exactly the Beveridge curve shifted, including data on inflows and outflows from unemployment, mismatch, and on-the-job search.

■ *Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views Federal Reserve Bank of Chicago, Federal Reserve Board of Governors, or of any of the other institutions with which the authors are affiliated.*

## References

- Ball, Laurence M., Daniel Leigh, and Prachi Mishra.** 2022. "Understanding U.S. Inflation during the COVID Era." NBER Working Paper 30613.
- Barlevy, Gadi.** 2002. "The Sullyng Effect of Recessions." *Review of Economic Studies* 69 (1): 65–96.
- Bernanke, Ben S.** 2012. "Recent Developments in the Labor Market." Speech, National Association for Business Economics Annual Conference, Arlington, Virginia, March 6. <https://www.federalreserve.gov/newsevents/speech/bernanke20120326a.htm>.
- Bernanke, Ben, and Olivier Blanchard.** 2023. "What Caused the U.S. Pandemic-Era Inflation?" Brookings Institution Hutchins Working Paper 86.
- Beveridge, William.** 1944. *Full Employment in A Free Society*. Crows Nest, Australia: Allen and Unwin.
- Blanchard, Olivier, Alex Domash, and Lawrence Summers.** 2022. "Bad News for the Fed from the Beveridge Space." Peterson Institute for International Economics Policy Brief 22-7.
- Blanchard, Olivier, and Jordi Galí.** 2010. "Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment." *American Economic Journal: Macroeconomics* 2 (2): 1–30.
- Carrillo-Tudela, Carlos, Hermann Gartner, and Leo Kaas.** 2023. "Recruitment Policies, Job-Filling Rates, and Matching Efficiency." *Journal of the European Economic Association* 21 (6): 2413–59.
- Chodorow-Reich, Gabriel, John Coghlanese, and Loukas Karabarbounis.** 2019. "The Macro Effects of Unemployment Benefit Extensions: A Measurement Error Approach." *Quarterly Journal of Economics* 134 (1): 227–79.
- Crump, Richard K., Stefano Eusepi, Marc Giannoni, and Ayşegül Şahin.** 2019. "A Unified Approach to Measuring  $u^*$ ." *Brookings Papers on Economic Activity* 50 (1): 143–214.
- Crump, Richard K., Marc Giannoni, Stefano Eusepi, and Ayşegül Şahin.** 2022. "The Unemployment-Inflation Trade-Off Revisited: The Phillips Curve in COVID Times." NBER Working Paper 29785.
- Crust, Erin E., Kevin Lansing, and Nicolas Petrosky-Nadeau.** 2023. "Reducing Inflation along a Nonlinear Phillips Curve." FRBSF Economic Letter 2023-17.
- Daly, Mary C., Bart Hobijn, Ayşegül Şahin, and Robert G. Valletta.** 2012. "A Search and Matching Approach to Labor Markets: Did the Natural Rate of Unemployment Rise?" *Journal of Economic Perspectives* 26 (3): 3–26.

- Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger.** 2013. "The Establishment-Level Behavior of Vacancies and Hiring." *Quarterly Journal of Economics* 128 (2): 581–622.
- Economic Sciences Prize Committee of the Royal Swedish Academy of Sciences.** 2010. *Markets with Search Frictions*. Stockholm: Royal Swedish Academy of Sciences.
- Elsby, Michael W., Bart Hobijn, and Ayşegül Şahin.** 2010. "The Labor Market in the Great Recession." *Brookings Papers on Economic Activity* 41 (1): 1–48.
- Elsby, Michael W., Ryan Michaels, and David Ratner.** 2015. "The Beveridge Curve: A Survey." *Journal of Economic Literature* 53 (3): 571–630.
- Faccini, Renato, and Leonardo Melosi.** 2023. "Job-to-Job Mobility and Inflation." *Review of Economics and Statistics*. [https://doi.org/10.1162/rest\\_a\\_01312](https://doi.org/10.1162/rest_a_01312).
- Farber, Henry S., Jesse Rothstein, and Robert G. Valletta.** 2015. "The Effect of Extended Unemployment Insurance Benefits: Evidence from the 2012–2013 Phase-Out." *American Economic Review* 105 (5): 171–76.
- Farber, Henry S., and Robert G. Valletta.** 2015. "Do Extended Unemployment Benefits Lengthen Unemployment Spells? Evidence from Recent Cycles in the U.S. Labor Market." *Journal of Human Resources* 50 (4): 873–909.
- Figura, Andrew, and Chris Waller.** 2022. "What Does the Beveridge Curve Tell Us about the Likelihood of a Soft Landing?" *FEDS Notes*, July 29. [https://www.federalreserve.gov/econres/notes/feds-notes/what-does-the-beveridge-curve-tell-us-about-the-likelihood-of-a-soft-landing-20220729.html?trk=public\\_post\\_comment-text](https://www.federalreserve.gov/econres/notes/feds-notes/what-does-the-beveridge-curve-tell-us-about-the-likelihood-of-a-soft-landing-20220729.html?trk=public_post_comment-text).
- Furman, Jason, and Wilson Powell III.** 2021. "What Is the Best Measure of Labor Market Tightness?" *Petersen Institute for International Economics blog*, November 22. <https://www.piie.com/blogs/realtime-economic-issues-watch/what-best-measure-labor-market-tightness>.
- Galí, Jordi.** 2011. "The Return of the Wage Phillips Curve." *Journal of the European Economic Association* 9 (3): 436–61.
- Galí, Jordi.** 2015. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. 2nd ed. Princeton, NJ: Princeton University Press.
- Goldin, Claudia.** 2006. "The Quiet Revolution That Transformed Women's Employment, Education, and Family." *American Economic Review* 96 (2): 1–21.
- Hall, Robert E., and Marianna Kudlyak.** 2022. "The Unemployed with Jobs and without Jobs." *Labour Economics* 79: 102244.
- Hershbein, Brad, and Lisa B. Kahn.** 2018. "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings." *American Economic Review* 108 (7): 1737–72.
- Kocherlakota, Narayana R.** 2010. "Inside the FOMC." *Speech, Federal Reserve Bank of Minneapolis, Marquette, Michigan, August 17, 2010*.
- Modestino, Alicia Sasser, Daneil Shoag, and Joshua Ballance.** 2020. "Upskilling: Do Employers Demand Greater Skill When Workers Are Plentiful?" *Review of Economics and Statistics* 102 (4): 793–805.
- Mortensen, Dale T., and Christopher A. Pissarides.** 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies* 61 (3): 397–415.
- Moscarini, Giuseppe, and Fabien Postel-Vinay.** 2023. "The Job Ladder: Inflation vs. Reallocation." *NBER Working Paper* 31466.
- Mukoyama, Toshihiko, Christina Patterson, and Ayşegül Şahin.** 2018. "Job Search Behavior over the Business Cycle." *American Economic Journal: Macroeconomics* 10 (1): 190–215.
- Petrongolo, Barbara, and Christopher A. Pissarides.** 2001. "Looking into the Black Box: A Survey of the Matching Function." *Journal of Economic Literature* 39 (2): 390–431.
- Petrofsky-Nadeau, Nicolas, and Lu Zhang.** 2021. "Unemployment Crises." *Journal of Monetary Economics* 117: 335–53.
- Phillips, A. William.** 1958. "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957." *Economica* 25 (100): 283–99.
- Pissarides, Christopher A.** 2000. *Equilibrium Unemployment Theory*. 2nd ed. Cambridge, MA: MIT Press.
- Şahin, Ayşegül, Jae Song, Giorgio Topa, and Gianluca Violante.** 2014. "Mismatch Unemployment." *American Economic Review* 104 (11): 3529–64.
- Shimer, Robert.** 1998. "Why Is the US Unemployment Rate So Much Lower?" In *NBER Macroeconomics Annual*, Vol. 13, edited by Martin Eichenbaum, Erik Hurst, and Ramey, 11–61. Cambridge, MA: MIT Press.
- Shimer, Robert.** 2012. "Reassessing the Ins and Outs of Unemployment." *Review of Economic Dynamics* 15 (2): 127–48.

- Weir, David R.** 1992. "A Century of U.S. Unemployment, 1890–1990: Revised Estimates and Evidence for Stabilization." *Research in Economic History* 14: 301–46.
- Woodford, Michael.** 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton University Press.
- Yashiv, Eran.** 2016. "Capital Values and Job Values." *Review of Economic Dynamics* 19: 190–209.

## Perspectives on the Labor Share

Loukas Karabarbounis

**T**he “labor share” refers to the fraction of an economy’s income that accrues to workers in exchange for their labor services. The evolution of the labor share has always fascinated, and often puzzled, economists, partly because it seems central to the question of how an economy is benefiting those who receive their income from working as opposed to those who receive it from owning capital. Ricardo (1817) famously argued that the distribution of income between labor and capital is the principal problem of political economy. Keynes (1939) noticed the United Kingdom and the United States had a stable labor share between the 1910s and the 1930s and claimed this was “a bit of a miracle.” Kaldor (1957) argued that the stability of the labor share is an important stylized fact of economic growth, which implied that an economy’s aggregate production could be written as a function of labor and capital in a convenient way. However, Solow (1958) was skeptical, arguing that this constancy “may be an optical illusion” and that we should not view the labor share as a fundamental constant of economics.

In 2022, the labor share in the United States was at its lowest level since the Great Depression. The headline estimate for the United States is a roughly 5 percentage points decline of the labor share between 1929 and 2022. The decline after World War II is even larger, at around 7 percentage points. The great majority

■ *Loukas Karabarbounis is Professor of Economics, University of Minnesota, and Research Consultant, Federal Reserve Bank of Minneapolis, both in Minneapolis, Minnesota. His email address is [loukas@umn.edu](mailto:loukas@umn.edu).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.2.107>.

of US industries exhibited labor share declines in recent decades. The United States is not unique, as we observe labor share declines in most countries of Europe and Asia and in emerging markets. The headline estimate for the world is a decline of roughly 6 percentage points in the labor share since 1980.

Conceptually, it may seem easy to divide income produced by an economy into labor compensation for providing labor services to production, and capital compensation for providing capital services to production. However, the measurement of the labor share is contentious because statistical agencies do not necessarily report the division of income between factors as economists conceptualize it. I lay out key issues in measuring the labor share, present various alternatives, and clarify some pitfalls in measuring practices.

In explaining why the labor share declined, I find it useful to consider five categories: technology, product markets, labor markets, capital markets, and globalization. The factors that have contributed to the labor share decline are intertwined, so no neat decomposition is possible without further analyses. My view is that the most plausible causes have technological origin. Developments such as the information age and automation, manifesting through changes in the cost of capital and the structure of product markets, caused the labor share to decline. However, this view is not necessarily dominant among economists, so I explain the mechanisms and evaluate the plausibility of each factor in light of available observations. I also highlight how the trend of the labor share relates to other macroeconomic trends and the implications that these trends have for economics research.

I conclude by discussing welfare, distributional, and policy implications of the labor share decline. Here the discussion is necessarily more speculative, because the consequences of the labor share decline partly depend on whether it will continue.

## **Measuring the Labor Share**

The measurement of the labor share is complicated by ambiguities about what constitutes labor versus capital income in the reports prepared by statistical agencies.<sup>1</sup> Should national or domestic income be used in the denominator of the labor share? How should we treat taxes? How to split proprietors' income between labor and capital? Should we include the government in our measurement of the labor share and what is capital income in the government sector? Should we include the imputed income that homeowners receive from living rent-free in their houses as capital income? How should we treat the depreciation of physical capital? Should income be netted out of measured investment expenditures? Or, should income be augmented to include some of the expenses that are currently missing from measured income but generate future returns?

<sup>1</sup>Influential predecessors for some of the issues that I discuss include Cooley and Prescott (1995), Gollin (2002), and Gomme, Ravikumar, and Rupert (2011).

Table 1

**The Distribution of Income in National Income and Product Accounts**

NIPA Item	Variable	Value (billions of dollars)
1.12 – 1	Net national income	18,281
1.10 – 21	Consumption of fixed capital	3,480
Equals	Gross national income	21,761
1.1.5 – 1	Gross domestic product	21,521
1.12 – 1	Net national income	18,281
1.12 – 2	Compensation to employees	11,448
1.12 – 9	Proprietors' income	1,554
1.12 – 12	Rental income	685
1.12 – 13	Corporate profits	2,470
1.12 – 18	Interest	515
1.12 – 19	Taxes on production and imports	1,553
1.12 – 20	Less: Subsidies	73
1.12 – 21	Business current transfer payments	161
1.12 – 25	Surplus of government enterprises	-12

*Source:* National Income and Product Accounts tables.

### Allocating Income between Labor and Capital

To illustrate the implications of these kinds of questions, Table 1 presents an example from the National Income and Product Accounts (NIPA) published by the US Bureau of Economic Analysis (BEA) for 2019.

The top part of the table concerns the denominator of the labor share calculation. National income net of depreciation is roughly \$18 trillion. Adding depreciation, which includes the consumption of both private and government fixed assets, yields a gross national income of close to \$22 trillion. The fourth line presents GDP. Products and incomes are conceptually the same objects, with differences arising because products are measured from the market value of new goods and services, while incomes are measured from compensations to factors and taxes on production. On the other hand, the distinction between national and domestic concepts is not merely statistical. “Domestic” refers to the income produced within a country using both domestic and foreign factors of production. “National” refers to the income produced by citizens of a country, irrespective of whether the income is produced domestically or abroad.

The second part of the table decomposes net national income into its components. The largest component is compensation to employees, which unambiguously belongs to labor income. The measure of compensation used by the Bureau of Economic Analysis includes wages and salaries, commissions, bonuses, tips, severance payments, supplements, in-kind benefits, employer contributions to pensions

and social insurance, and exercised “nonqualified” stock options (that is, options on which a taxpayer pays ordinary income taxes on the gain between their grant price and exercise price). Corporate profits, rentals, interest income, business transfers, and the surplus of government enterprises unambiguously are compensation to capital.<sup>2</sup>

This leaves us with the choice of how to allocate taxes (less subsidies) on production and the income of proprietors. Taxes on production include indirect taxes such as sales and excise taxes, property taxes, motor vehicle licensing fees, and import duties. They do not include direct income taxes, which are included either in corporate profits or in compensation to employees. Examples of subsidies are those that accrue to farmers and local governments. A reasonable approach is to allocate taxes less subsidies proportionally between labor and capital. This avoids tilting the balance between unambiguous labor and capital income.<sup>3</sup>

Proprietors’ income includes both labor and capital components. As an example, consider an Uber driver. The Internal Revenue Service observes only gross receipts less expenses occurred while driving, not the split of income between labor and capital. The driver earns labor income in compensation for time spent driving passengers to destinations. However, the driver also earns capital income, because they operate their car to earn income. Thus, part of the total income of the Uber driver includes the depreciation cost and the opportunity cost of using their car instead of earning rent from leasing it to another driver or interest from investing in a financial asset.

There are several ways to handle proprietors’ income. The most unsatisfactory way is to exclude it from the numerator of the labor share, because this amounts to treating proprietors’ income entirely as capital income. A more satisfactory alternative is to impute the labor income of proprietors by multiplying employees’ average wage with proprietors’ hours worked. Another satisfactory alternative is to assume that proprietors use labor as intensively as the rest of the economy does. Considered in terms of our Uber driver example, the first alternative assumes that the hourly wage of an Uber driver equals the wage of drivers employed by driving companies in the rest of the economy, whereas the second alternative assumes that the share of labor income of the Uber driver is equal to the labor share of driving companies who employ drivers.

For my analysis of the United States, I construct three measures of the labor share. The first, called “Total: Proprietors Same Share,” covers the entire economy and allocates taxes on production and proprietors’ income proportionally

<sup>2</sup>Online Appendix A discusses measurement difficulties for compensation of employees arising from equity-based compensation (Eisfeldt, Falato, and Xiaolan 2023), for corporate profits arising from profit shifting (Güvenen et al. 2022), and for rentals arising from imputations of owner-occupied housing services. It also discusses how tax incentives may shift reported labor and capital income within the corporate sector and the treatment of corporate housing. Finally, it documents differences between my measures and the labor share measure produced by the US Bureau of Labor Statistics.

<sup>3</sup>Allocating ambiguous income proportionally between labor and capital is equivalent to excluding it from the denominator:  $\text{Labor Share} = \text{Labor Income} / (\text{Total Income} - \text{Ambiguous Income})$ .



between labor and capital. It equals compensation to employees divided by GDP less taxes on production and proprietors' income. The second measure, called "Total: Proprietors Same Wage," continues to allocate taxes on production proportionally between labor and capital but assumes equal earnings between proprietors and employees for imputing the labor income of proprietors. It equals compensation to employees divided by GDP less taxes on production, scaled by a time-varying factor that equals one plus the ratio of proprietors' to employees' labor input.<sup>4</sup>

The third measure is the corporate labor share, defined as compensation to employees in the corporate sector divided by corporate value added less corporate taxes on production. For the United States, the benefit of focusing on the corporate sector is that proprietors are not incorporated, and there is no ambiguity about how to split their income. The corporate sector accounts for roughly 50 to 60 percent of GDP. It excludes housing, government, and unincorporated businesses.

For analyses of international data, the System of National Accounts (SNA) does not have a separate entry for proprietors' income. The closest alternative is to proxy for proprietors' income with the operating surplus of private unincorporated businesses (mixed income). The caveat is that some private businesses may be paying wages to their owners, and some proprietors may be legally incorporated. I believe the corporate labor share offers a balanced alternative in terms of availability, comparability, and measurement concerns. In terms of availability, Karabarbounis and Neiman (2014b) compiled data for a large cross section of countries from the SNA. Focusing on the corporate sector has the advantage of holding the sectoral composition of economic activity relatively constant across countries, as the cross-country variation in industry and sectoral shares is much larger than the variation observed within a country.

### **The Official BLS Measure of the Labor Share**

The US Bureau of Labor Statistics (BLS) produces the official measure of the labor share in the United States (at <https://rb.gy/fud0l>). The BLS measure is most similar to "Total: Proprietors Same Wage," because it adjusts for proprietors' income assuming the same wage between employees and proprietors. One difference of the BLS measure is that it only covers the nonfarm business sector, which accounts for roughly three-quarters of GDP as it excludes government and owner-occupied housing services.

The labor share "Total: Proprietors Same Share" is preferred to "Total: Proprietors Same Wage," because I view the assumption of equal factor shares across businesses' and proprietors' technologies as more natural than the assumption that proprietors earn the same hourly wage as employees. In fact, Hurst and Pugsley (2011) document significant nonpecuniary benefits from self-employment, so

<sup>4</sup>To calculate this labor share, define Total Income as Capital Income + Compensation of Employees + Proprietors' Income. Assuming that  $w$  is the return to labor for both the employees and the proprietors, we impute Proprietors' Labor Income =  $w \times$  Proprietors' Labor Input. Then, Labor Share = (Compensation of Employees/Total Income)  $\times$  (1 + Proprietors' Labor Input/Employees' Labor Input).

assuming equal wages has the unappealing implication that proprietors are better off than employees. By contrast, I do not see why production technologies should change upon crossing the legal boundaries that separate organized businesses from proprietorships.

There are two arguments why one might want to exclude housing from the labor share. First, owner-occupied rental income may be imputed with significant measurement error (Rognlie 2015). Second, housing is a broadly held asset whose returns accrue also to workers, and thus the split between capital and labor commonly associated with the distribution of income does not apply to owner-occupied housing. I do not view this as a problem, because there is no presumption that laborers and capitalists are two mutually exclusive groups of households. Rather, the labor share may be a poor proxy for inequality, given that dividends and interest income may also accrue to workers. Almost all of the value added of the housing sector is capital income, so including it in the measurement decreases the labor share relative to what it would be if the housing sector were excluded.

The most common argument for excluding government is that including it biases the labor share upward, because households do not consume capital income generated from government's production. I do not find this argument compelling, because even if the government does not produce direct payments of capital income, government's compensation to employees is still informative about distribution. Further, because the government is the largest industry by value added, shifts in its size could affect the aggregate labor share. The most compelling argument for excluding government from the labor share calculation, at least for certain research purposes, is that its optimization decision is quite different from the optimization decision of the private sector. If one uses the labor share to learn about production technologies that are outcomes of private decisions, then it might be reasonable to exclude government from the measurement of the labor share.

### **Additional Measurement Issues**

The appropriateness of any measure of the labor share will ultimately depend on the application for which it is being used. Robust scientific practice is to examine the sensitivity of the conclusions to alternative measurements and to understand where differences might arise. The discussion has already mentioned a number of different ways of measuring the labor share, such as excluding the government sector or owner-occupied housing, or different ways of dividing the proprietors income between labor and capital. Here, I list some other choices that can arise in calculating the labor share.

*National or Domestic Product in the Denominator of the Labor Share?* For analysis of the distribution of income across citizens of a country, the appropriate concept of income is the national one, because some citizens work or invest abroad. For analysis of production, factor supply and demand, and productivity of a country, the appropriate concept of income is the domestic one, because domestic technologies and institutions affect them. For the measurement of the US labor share, this distinction is not important, as the ratio of national to domestic income is quite

stable around one. I have used domestic product concepts, because they more closely correspond to industry value-added concepts and are more readily available for foreign countries.

*Treatment of Depreciation.* All measures of the labor share have gross output in the denominator, and thus include the depreciation of capital goods arising from wear, tear, and obsolescence. Aggregate depreciation rates have been increasing over time, as economies have shifted their composition of fixed assets toward higher-depreciating assets such as software and computers. Mechanically, this compositional change tends to lower the labor share.

Some research that emphasizes the distributional role of the labor share chooses to exclude depreciation from the measure of income in the denominator, because it does not represent consumption by households. From a production perspective, gross product is the more appropriate notion of income, because depreciation absorbs resources. However, depreciation is imputed by statistical agencies, and thus net product is subject to larger measurement errors than gross product. Additionally, the labor share of income net of depreciation measures the net income of capital owners relative to that of workers at a point in time. Net incomes at a point in time are not very informative about inequality outside of steady state, because capital owners' welfare is the present discounted value of consumption flows and not current consumption. For perspectives on depreciation and the labor share, useful starting points are Karabarbounis and Neiman (2014a), Piketty (2014), and Rognlie (2015).

*Treatment of Investments.* Some measures of the labor share exclude part of or the entire investment expenditure of an economy from GDP in the denominator when calculating the labor share. There are two ways to rationalize this practice. First, investment expenditures are not consumed, and therefore this alternative measure is more directly related to inequality between owners of labor and owners of capital. Second, this alternative measure is more robust to measurement error in investments when evaluating the drivers of the labor share.

This practice causes confusion, so let me discuss with an example. The economy produces \$100 of output. On the expenditure side, households purchase and consume \$80 of goods and services, and corporations purchase \$20 of capital goods to use in future production. On the income side, corporations pay \$60 to workers and earn \$40 of (accounting) profits. Out of these \$40, \$20 are paid to households in the form of dividends, and \$20 are used to purchase capital goods. The standard measure of the labor share is 0.6. An argument for excluding investment from income, thus increasing the labor share to 0.75, is that we allocate the remaining \$80 between labor income (the "laborers") and dividends (the "capitalists"), which both are consumed. I do not find this argument persuasive. The \$20 of investment produce future returns to owners of capital. It is not economically reasonable to assume that changes in corporate valuations do not matter for welfare. For example, capitalists optimally postpone consumption when the productivity of investment or the demand for goods from the rest of the world is higher today relative to the

future. Capitalists' welfare is the present discounted value of consumption flows, not consumption in the current period.

I recognize the difficulty of measuring some forms of investments, such as intellectual property products, especially in the early years of the national income and product accounts. But there are measurement difficulties with other NIPA items, such as proprietors' labor input and income. Further, the labor share that excludes investment is not informative about aggregate production, unless one thinks that corporate financial and payout policies have much to do with production. Returning to our example, if corporations increased their dividend payout to \$30 and reduced investment to \$10, the labor share excluding investment would increase, but production would not change. Taking measurement at face value, the Bureau of Economic Analysis appropriately treats expenses on intellectual property products as part of output to the extent that these expenses are not consumed within a period and produce future returns to their owners. To the extent that these investments augment capital owned by businesses, the BEA appropriately treats the income generated by this capital stock as capital income. For alternative views on these issues, useful starting points include Atkeson (2020), Koh, Santaaulàlia-Llopis, and Zheng (2020), and Barro (2021). Online Appendix A discusses how unmeasured intangibles and durable services affect the measurement of the labor share.

## Observations on the Labor Share Decline

This section begins with analyses of the US labor share and then proceeds with analyses for a large cross-section of other countries.

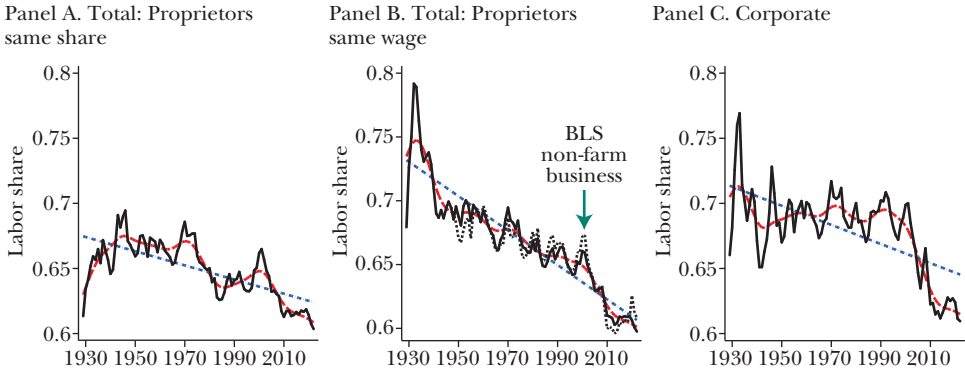
### United States

*Aggregate Labor Shares.* Figure 1 presents the evolution of the three labor share series in the United States between 1929 and 2022.<sup>5</sup> For each labor share series, I plot the linear trend with the dotted blue line and, to visualize nonlinearities, the Hodrick–Prescott trend with the red long-dashed line. Common issues in the labor share literature are that starting and ending points matter quite a bit for estimating trends and that trends appear to be nonlinear, especially in samples that extend to years before World War II. For these reasons, the upper panel of Table 2 presents summary statistics during four different subperiods, a subperiod that covers the years before World War II and three subperiods after World War II, all of roughly the same length.

Beginning with the left panel of Figure 1, the labor share measure “Total: Proprietors Same Share” increases during roughly the first 15 years of the sample and then declines between the end of World War II and the end of the sample. Table 2 shows that the labor share increases by 3 percentage points per decade between 1929 and

<sup>5</sup>Online Appendix B presents three additional measures of the labor share, that either exclude housing, or exclude the government, or add government's imputed capital income to GDP.

Figure 1  
**Labor Share in the United States, 1929–2022**



Source: Author’s calculations available in Karabarbounis (2024).  
 Note: Solid black line is the labor share measure, dotted blue line is the linear trend of each measure, and red long-dashed line is the Hodrick-Prescott trend of each measure with a smoothing parameter of 100. The short-dashed black line is the BLS measure of the labor share for the nonfarm business sector, scaled to equal “Total: Proprietors Same Wage” in the first year of its observation.

Table 2  
**Trends of the US Labor Share**

		Total: same share	Total: same wage	Corporate
Subperiod (percentage points per decade)	1929–1945	3.0	–3.9	–3.2
	1946–1970	–0.2	–0.9	–0.1
	1971–1995	–1.6	–1.1	–0.1
	1996–2022	–1.8	–2.3	–3.3
Period: 1929–2022 (percentage points)	Linear trend	–5.1	–12.5	–6.8
	HP(100) trend	–2.1	–13.4	–9.0
	MA(5) trend	–2.1	–11.7	–7.4
Period: 1946–2022 (percentage points)	Linear trend	–6.3	–9.4	–7.4
	HP(100) trend	–6.6	–9.1	–7.0
	MA(5) trend	–7.3	–9.0	–8.2

Source: Author’s calculations available in Karabarbounis (2024).  
 Note: In the first panel, entries are percentage points changes per decade in different measures of the labor share. The estimates are from separate regressions of labor share measures on a linear trend within each subperiod. In the second panel for the period between 1929 and 2022, entries are estimates of the trend of each labor share measure using a linear trend, a Hodrick–Prescott trend with a smoothing parameter of 100, and a five-year moving average. The third panel repeats these estimates for the period between 1946 and 2022.

1945 and then declines by 0.2 percentage point per decade between 1946 and 1970, by 1.6 percentage points per decade between 1971 and 1995, and by 1.8 percentage points per decade between 1996 and 2022.

The second panel of Figure 1 plots my series “Total: Proprietors Same Wage” alongside the official labor share series from the Bureau of Labor Statistics. My series tracks the BLS series extremely well, with a correlation of 0.96. The labor share “Total: Proprietors Same Wage” exhibits the largest decline among all measures. The upper panel of Table 2 shows that the main difference with the other measures is in the first subperiod, when the labor share declines by 3.9 percentage points. I discuss measurement difficulties with proprietors’ labor input in online Appendix A, because they might be partly responsible for the large decline in the first subperiod.

In the third panel of Figure 1, we see that the corporate labor share behaves quite differently than the other two measures. As the upper panel of Table 2 shows, the corporate labor share declines substantially before World War II. Further, the corporate labor share is more stable than the other measures in the first decades after the World War II and declines by more than the other measures during the last subperiod of the sample.

The lower two panels of Table 2 summarize trends of the labor share for the whole sample (1929 to 2022) and the sample after World War II (1946 to 2022). I use three different ways to estimate the trend: a linear trend, a Hodrick–Prescott trend with a smoothing parameter of 100, and a five-year moving average. For “Total: Proprietors Same Share,” the linear trend shows a decline of 5.1 percentage points in the whole sample. The nonlinear trends display a smaller decline at 2.1 percentage points. The sensitivity to applying a linear trend disappears in the sample after World War II, with the labor share declining between 6 and 7 percentage points. The “Total: Proprietors Same Wage” measure exhibits the largest changes among all measures, declining by roughly 12 percentage points in the whole sample and by 9 percentage points in the sample after World War II. The corporate labor share declines by roughly 8 percentage points beginning in both samples.

In sum, all of these labor share series experience declines, although to a varying degree and sometimes at different times. A baseline estimate averages the trends from the “Total: Proprietors Same Share” and the corporate labor share measure, which are my preferred measures, to conclude that the labor share declined by 5.4 percentage points when starting in 1929 and by 7.1 percentage points when starting in 1946. However, the two measures behave quite differently. Most of the difference is compositional, because the corporate sector has a higher labor share and its contribution to GDP increased by almost 10 percentage points between the beginning of the sample and 1970. The recent decline of the labor share is more similar between these two measures.

*Industry Labor Shares.* The integrated BLS–BEA industry production accounts (US KLEMS) combine output and intermediate inputs data from the Bureau of Economic Analysis industry accounts with measures of factor inputs and compensations from the Bureau of Labor Statistics productivity program. Compensation to

employees in the integrated accounts adjusts the original BEA compensation to employees data in order to account for proprietors' labor income (Fleck et al. 2014). Following the BLS practice, the adjustment assumes equal wages among employees and proprietors conditional on demographics.

Value added at the industry level from the BLS–BEA integrated accounts differs significantly from value added in the Bureau of Economic Analysis's industry accounts, with gaps ranging from roughly –20 to 40 percent. Given these differences, I consider two labor share measures at the industry level. Both use compensation from the integrated BEA–BLS accounts in the numerator, because it adds proprietors' labor income. The first one, which I call “BLS Industry,” divides compensation to employees by value added for each industry from the BEA–BLS integrated accounts. The second measure, which I call “BEA Industry,” instead uses the levels of value added and taxes on production from the BEA industry accounts to impute value-added net of taxes on production in the integrated accounts.

Table 3 presents summary statistics across US industries for the two measures of the labor share. BLS Industry is used in the first three columns, and BEA Industry is used in the remaining three columns. The first and fourth columns show the value-added share of each industry in total value added, averaged over the entire sample. The value-added shares are very similar across the two measures.

The second and fifth columns present the average labor share of each industry over the entire sample. The BEA Industry labor share allocates taxes on production proportionally between labor and capital income, and so it is higher than the BLS Industry labor share. Industries with low labor share include mining and oil, utilities, and real estate. High labor share industries are construction and services such as management, health, and education. The patterns are very similar across measures, with the main exception being the government. The value-added share of the government under the Bureau of Labor Statistics measure is significantly larger than the value-added share under the Bureau of Economic Analysis measure, because the integrated accounts impute capital income for the government that exceeds depreciation. For the same reason, the labor share of the government under the BLS measure is significantly lower than the labor share under the BEA measure.

The third and sixth columns present the trend of the labor share by industry. Given the relatively short sample, and the fact that choice of endpoints can clearly make a difference, as demonstrated in the earlier discussion, I calculate the trend as the difference in the average labor share between the second half and the first half of the sample. The trends are similar between the two labor share measures. Out of 20 industries, only agriculture, utilities, arts and entertainment, and other services exhibit increases in their labor shares, whereas the remaining 16 industries exhibit declines of their labor shares. The median decline is around 5 percentage points, and the average decline weighted by industry share in value added is around 6 percentage points. The magnitude of these declines matches well with the 5.5 percentage points decline after 1987 in the measure “Total: Proprietors Same Wage” for the aggregate economy, which is closest to the official Bureau of

Table 3

**Statistics of US Industries, 1987–2021**

Industry	BLS industry			BEA industry		
	Value added	Labor share		Value added	Labor share	
	Share	Level	Trend	Share	Level	Trend
Agriculture	1.2	63.8	2.2	1.3	66.5	5.8
Mining, oil	1.4	33.5	−13.9	1.4	36.7	−15.1
Utilities	1.7	28.6	1.1	1.6	32.2	1.5
Construction	4.1	83.2	−8.7	4.4	84.8	−9.1
Manufacturing	13.8	55.9	−11.6	14.5	57.2	−11.6
Wholesale	4.7	65.9	−14.3	5.0	67.2	−16.7
Retail	5.1	78.2	−7.1	5.4	79.4	−5.1
Transportation	2.9	70.6	−8.4	3.0	72.6	−9.1
Information	4.6	42.8	−7.3	4.9	43.7	−8.3
Finance	6.8	58.6	−3.1	7.2	59.9	−3.1
Real estate	12.5	7.8	−1.0	11.9	8.9	−1.1
Professional, business	6.4	81.7	−6.1	6.8	82.5	−6.5
Management	1.6	87.4	−1.0	1.7	89.8	−1.6
Administrative	2.6	83.0	−5.2	2.8	83.7	−5.5
Education	1.0	85.1	−6.4	1.1	88.3	−7.1
Health	6.5	89.2	−5.1	6.9	90.6	−5.6
Arts, entertainment	0.9	81.3	1.4	0.9	83.6	1.8
Accommodation, food	2.4	74.6	−3.6	2.6	75.6	−3.6
Other services	2.3	82.0	8.0	2.5	83.7	8.5
Government	17.3	61.4	−7.1	14.1	81.2	−9.4
Median	3.5	72.6	−5.7	3.7	77.5	−5.6
Average	5.0	65.7	−4.9	5.0	68.4	−5.0
Weighted average		59.4	−6.1		64.2	−6.6

*Source:* Author's calculations from the Integrated Industry-Level Production Account (KLEMS) at <https://www.bea.gov/data/special-topics/integrated-industry-level-production-account-klems>, with details in Karabarbounis (2024).

*Note:* Value-added share and labor share level are averages by industry between 1987 and 2021. Labor share trend is the change of the average labor share in the period between 1987 and 2003 and the average labor share in the period between 2005 and 2021. For methodology, details, and additional discussion, see online Appendix A.

Labor Statistics measure of the labor share and thus the most comparable measure to the averaged industry measure.

Given the significant labor share declines observed for the great majority of industries, it is not surprising that the decline of the aggregate labor share reflects within-industry declines, rather than a reallocation of economic activity from high to low labor share industries. A formal shift-share decomposition shows that roughly



100 percent of the decline of the aggregate labor share is because of the within-industry component, and not because of shifts between industries.<sup>6</sup>

## **World**

To calculate the labor share for a range of countries, I use data from the Penn World Tables (Feenstra, Inklaar, and Timmer 2015, PWT version 10.01). This database is compiled from national accounts from 1950 onward with information on output, inputs, productivity, and labor shares. My main series is the labor share that allocates mixed income and taxes on production proportionally between labor and capital, thus resembling the “Total: Proprietors Same Share” measure (see online Appendix A for more details).

Figure 2 presents the labor share of the 16 largest economies of the world based on 2015 GDP. Out of these, we can see 13 economies whose labor share has declined. The pattern of these declines is not related in any obvious way to geography, level of the labor share, or level of development. We observe labor-share declines in advanced Anglo-Saxon economies (Australia, Canada, and United States), advanced European economies (France, Germany, Italy, and Spain), advanced Asian economies (Japan and Korea), and emerging markets (China, India, Mexico, and Thailand). The only countries with increases are Brazil, the United Kingdom, and Russia.

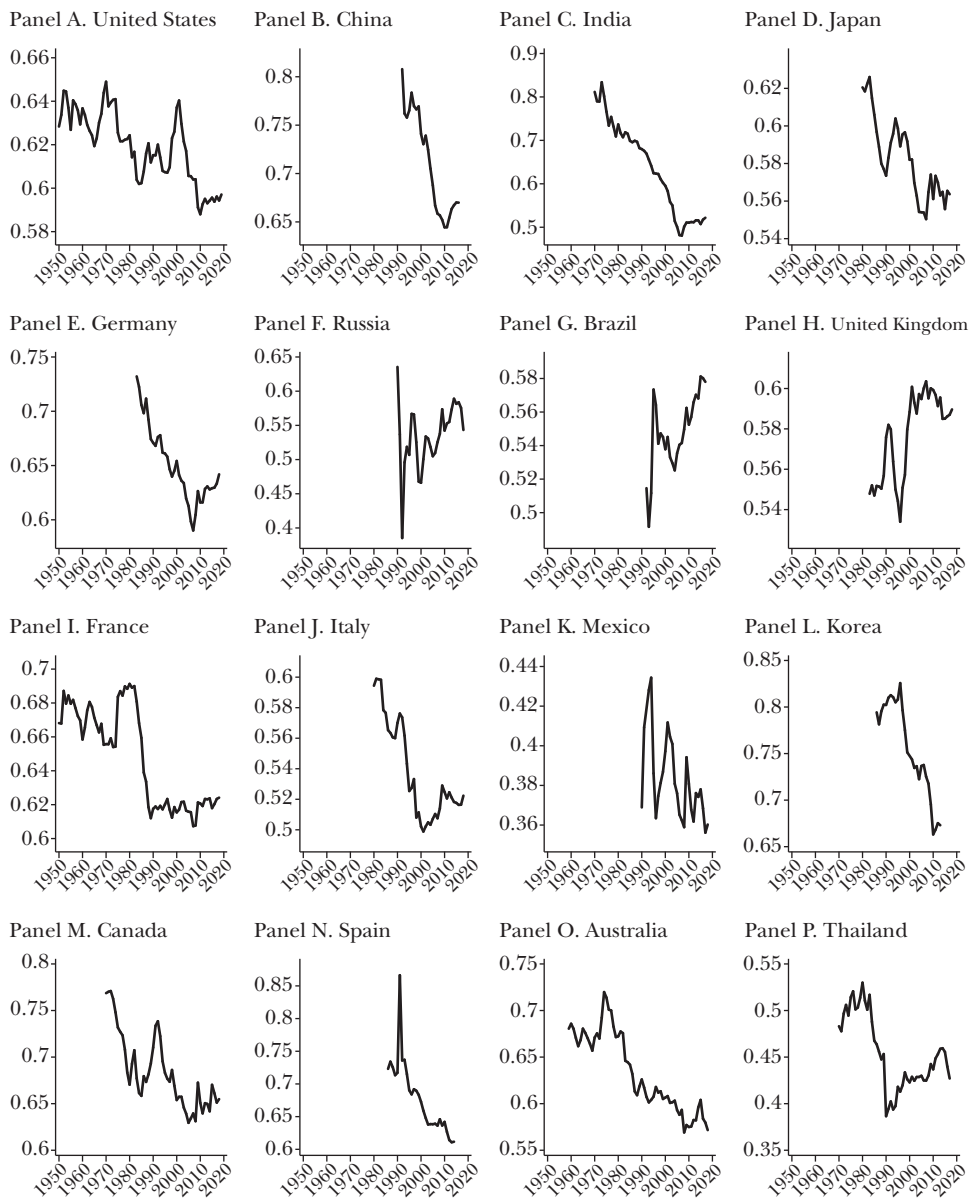
Table 4 presents the global patterns of the labor share decline more systematically. For each country and subperiod, I estimate the trend of the labor share using a linear trend, the change in the Hodrick–Prescott trend, and the changes in five-year moving averages. Each row of the table presents statistics of the distribution of trends using the various methods.

In the top panel, the sample between 1970 and 2019 covers 16 countries, which account for 39 percent of world GDP. Across the various detrending methods, the median country experienced a decline of roughly 8 to 11 percentage points since 1970, with an interquartile range between 4 and 13 percentage points. Weighted by GDP, the average decline is around 8 percentage points.

For the middle panel, the period covering 1980 to 2019 has 27 countries, which account for roughly half of world’s GDP. The median country experienced a decline between 6 and 7 percentage points. Weighted by GDP, the average decline is 5.9 percentage points. For the period between 1995 and 2019, we have 57 countries, which account for 85 percent of world GDP. In this shorter sample, both the median and the average country (weighted by GDP) experienced a decline of around 2 to 3 percentage points.

<sup>6</sup>This result confirms qualitatively previous analyses at the industry level such as Elsby, Hobijn, and Sahin (2013) for the United States and Karabarbounis and Neiman (2014b) for the world. I discuss quantitative differences in online Appendix A.

*Figure 2*  
**Labor Share around the World**



*Source:* Author's calculations based on Penn World Tables (Feenstra, Inklaar, and Timmer 2015, PWT version 10.01). Details in Karabarbounis (2024).

*Table 4*  
**Trends of the World Labor Share**

Period: 1970–2019	Linear trend	HP(100) trend	MA(5) trend	Countries	% GDP
Twenty-fifth percentile	−14.1	−13.7	−13.0		
Median	−10.5	−8.2	−8.9		
Seventy-fifth percentile	−7.3	−4.1	−4.5		
Average	−15.1	−13.2	−12.6		
Weighted average	−8.7	−7.8	−7.7	16	39
Period: 1980–2019	Linear trend	HP(100) trend	MA(5) trend	Countries	% GDP
Twenty-fifth percentile	−11.5	−9.6	−9.8		
Median	−5.9	−6.8	−7.1		
Seventy-fifth percentile	−2.2	−2.4	−2.4		
Average	−7.4	−7.3	−7.2		
Weighted average	−6.2	−5.9	−5.9	27	52
Period: 1995–2019	Linear trend	HP(100) trend	MA(5) trend	Countries	% GDP
Twenty-fifth percentile	−6.0	−4.8	−4.9		
Median	−2.4	−2.6	−2.6		
Seventy-fifth percentile	1.6	0.8	0.9		
Average	−3.7	−3.6	−3.5		
Weighted average	−3.2	−2.6	−2.4	57	85

*Source:* Author’s calculations based on Penn World Tables (Feenstra, Inklaar, and Timmer 2015, PWT version 10.01). Details in Karabarbounis (2024).

*Note:* Entries are the trend of the labor share using a linear trend, a Hodrick–Prescott trend with a smoothing parameter of 100 and a 5-year moving average. The last two columns present the number of countries and their share of world GDP in each sample period.

## **Causes of the Labor Share Decline**

While the decline in the labor share varies in magnitude and timing across countries and industries, the systematic pattern of a decline suggests that common factors may underlie the decline. This often overlooked observation restricts significantly the freedom of researchers to propose theories that account for the decline of the labor share. It may well be that idiosyncratic factors have higher explanatory power than common factors in some countries or industries. Nonetheless, it is unlikely that countries as different as India and the United States or as Korea and France all experienced changes in labor share for unrelated reasons. Similarly, the systematic pattern of a decline across industries also argues against explanations based on compositional factors.

I discuss the factors that might have caused the decline of the labor share using the profit-maximization problem of a firm that operates in partial equilibrium and chooses prices and quantities of output and inputs. The production function describes how output  $y$  is produced with physical capital  $k$  and labor  $\ell$ . It features constant returns to scale, a constant elasticity of substitution,  $\sigma$ , between

the two inputs and a distribution factor,  $\alpha$ , which weights the importance of capital in production. It also features technology  $A_k$ , which augments capital, and technology  $A_\ell$ , which augments labor. The firm faces product demand  $p(y)$  and labor supply  $\ell(w)$ , and internalizes that its price  $p$  and its wage  $w$  depend on output and labor. The firm takes as given the (user) cost of capital  $R$ .

This leads to the following solution for the labor share of income,

$$s_\ell = \frac{w\ell}{py} = \left(\frac{1}{\mu\theta}\right) \left[1 - \alpha^\sigma \left(\frac{A_k}{\mu R}\right)^{\sigma-1}\right],$$

where  $\mu \geq 1$  is the markup of price over marginal cost of production, which distorts all input choices, and  $\theta \geq 1$  is the markup of the marginal revenue product of labor over the wage, which distorts only labor. Given an elasticity of substitution  $\sigma$  and a distribution factor  $\alpha$ , the labor share depends on four factors: (1) capital-augmenting technology  $A_k$ ; (2) cost of capital  $R$ ; (3) product market markup  $\mu$ ; and (4) labor market markup  $\theta$ . The labor share does not depend on labor-augmenting technology  $A_\ell$  and the wage  $w$ . Online Appendix C explains the reasoning, provides the derivations, considers the case without constant returns to scale, and discusses how this solution might also apply in general equilibrium.

### Technology

Perhaps the most intuitive explanation for the decline of the labor share is that capital-augmenting technology has increased over time. Using our labor share solution, we see that this requires an increase in  $A_k$  and  $\sigma > 1$ . The economics of why parameter  $\sigma$  affects the relationship between the labor share and capital-augmenting technology are fairly straightforward. Factors are substitutes when  $\sigma > 1$ , and thus, following an increase in capital-augmenting technology, production requires more capital relative to labor to accomplish the same level of output. In this case, the labor share decreases.<sup>7</sup>

Despite the appealing economics of how capital-augmenting technology might have affected the labor share, this is also the most irrefutable explanation of the labor share decline. In order to rationalize the decline of the labor share, one could claim that  $A_k$  decreases over time and factors are complements,  $\sigma < 1$ . While this case is nonintuitive because we suspect that technology is improving over time, there is little hope of differentiating between these two explanations. The fundamental problem is that factor-augmenting technologies are not easy to conceptualize and measure. For example, are robots physical capital, or do they augment high-skilled labor's technology? Do improvements in an organization's management increase or decrease the productivity of capital relative to labor?

<sup>7</sup>The elasticity of substitution is the percent increase in the capital-labor ratio  $k/\ell$  in response to a 1 percent increase in relative prices  $w/R$  (Robinson 1933). To give an example, if  $\sigma = 0$  (Leontief production), then  $A_k k = A_\ell \ell$ , so an increase in  $A_k$  necessitates a decrease in  $k/\ell$  and, for given  $R/w$ , generates an increase in the labor share.

The development of task-based models of the labor market has allowed economists to think more concretely about the role of technology for factor shares. These models emphasize the conceptual difference between tasks, which produce services of output, and skills, which are workers' capabilities for performing tasks. The key assumption is that some types of services can be produced with either capital or labor, whereas other services are produced only with labor. Automation decreases the number of tasks which are produced only with labor, enabling capital to substitute for labor in a larger share of tasks.

The economics of automation are similar to an increase in the distribution parameter  $\alpha$  in the production function of the firm. As our labor share solution shows, the increase in  $\alpha$  unambiguously reduces the labor share, for any value of  $\sigma$ . In fact, some researchers use the labor share as a proxy for the displacement of tasks due to automation (Acemoglu and Restrepo 2022). While the economics are similar, task-based frameworks are empirically more appealing than frameworks with factor-augmenting technologies because they are easier to refute using proxies for automation, such as the adoption of robots and the share of routine jobs. One difference between factor-augmenting technology and task-based frameworks is that in the former capital-augmenting technology increases wages, whereas in the latter automation could displace labor demand and lower wages (Acemoglu and Autor 2011).

### Cost of Capital

A closely related explanation for the decline of the labor share is that the cost of capital  $R$  has decreased over time and  $\sigma > 1$ . A decrease in the cost of capital relative to that of labor unambiguously increases demand for capital relative to that of labor. If the increase in the demand for capital is sufficiently strong to offset the decline in its cost (that is, when  $\sigma > 1$ ), payments to capital  $Rk$  increase relative to payments to labor  $w\ell$ , and the labor share declines. Thus, the economics of a decrease in the cost of capital are similar to the economics of capital-augmenting technology. Despite this similarity, explanations of the decline of the labor share based on a falling cost of capital are more empirically appealing. Unlike capital-augmenting technology, which is difficult to conceptualize and measure, we have a fairly good understanding of the determinants of the cost of capital and an easier time measuring its components.

What determines the cost of capital? Although a few firms rent their physical capital, most firms own their physical capital, and so by "cost of capital" economists mean the opportunity cost of using capital in the production process. To measure opportunity cost in this situation, it is useful to think about how a firm uses a unit of capital in production and, had it not used this unit, what it would have done with the freed-up resources (Hall and Jorgenson 1967). First, the firm could sell physical capital and invest in financial assets. Second, the firm incurs the cost of physical damage to capital and gains or losses from changes in the price of capital. Third, the firm pays sales taxes to purchase capital and further taxes on the income generated by capital. Fourth, the price of capital goods that the firm purchases may

change over time, relative to the price of its output. Fifth, the firm may be uncertain about its demand and productivity, and so it needs to be compensated for using physical capital instead of investing in safe assets. Finally, the firm may incur costs of adjusting its capital stock and of raising financing to purchase the capital.

Several components of the cost of capital have exhibited secular declines over time. The price of capital goods has declined relative to the price of consumption goods both in the United States (Greenwood, Hercowitz, and Krusell 1997) and for various other countries (Karabarbounis and Neiman 2014b). For example, according to the industry accounts from the Bureau of Economic Analysis, the (quality-adjusted) nominal price of computers and electronics declined by 52 percent between 2001 and 2021, and the price of motor vehicles increased by only 15 percent. By contrast, the price of food products increased by 59 percent, the price of housing services by 72 percent, and the price of educational services by 75 percent. The decline of the relative price of capital goods reflects technological factors that favor faster productivity growth in investment than in consumption-producing sectors and globalization, which decreased the cost of importing capital goods.

Up through 2021, real interest rates also generally declined in the past 40 years around the world. This decline reflects factors such as low aggregate productivity growth, the aging of the population, and credit market liberalization, as discussed by Henry (2003), Bernanke (2005), Eggertsson, Mehrotra, and Robbins (2019), and Mian, Straub, and Sufi (2021). Corporate income taxes also decreased for several countries. Kaymak and Schott (2023) connect the decline of corporate income taxes to the falling labor share of income.

The argument that the fall in the cost of capital caused a decline of the labor share requires an elasticity of substitution between capital and labor that is higher than one. It is fair to say that this elasticity is one of the most controversial parameters in economics. Earlier research concluded mostly that the elasticity of substitution is lower than one (Chirinko 2008), but some recent research has argued that it is higher than one (Karabarbounis and Neiman 2014b; Hubmer 2023).

There are two reasons to favor a high estimate of the elasticity of substitution. First, methodologies estimating low elasticities typically use wage data as an independent variable and omit variables correlated with wages in the residual of their regressions (as explored at greater length in online Appendix D). Second, low elasticities of substitution at the establishment level are not very informative about the aggregate elasticity of substitution in the long run. When the cost of capital decreases, new firms may enter with technologies that are more capital intensive than the technologies used by incumbents. Or, among incumbents, a larger share of economic activity may be concentrated at firms with relatively high capital shares. Such substitution patterns may cross local, industry, or even national boundaries, and thus it is not feasible to estimate the aggregate elasticity of substitution using micro-level data alone or variation within an industry or geographic area. To draw a parallel with the labor supply literature, aggregation theory demonstrates that individual-level estimates of the elasticity of labor supply may not be very informative about overall the economy's labor response to changes in productivity and taxes.

This is an often misunderstood point, so consider a concrete example that illustrates why firm-level technologies may not be useful in understanding the aggregate labor share. Consider an industry populated by two perfectly competitive firms. The labor-intensive firm produces one unit of output with one unit of labor. The capital-intensive firm produces  $A_k$  units of output with one unit of capital. By construction, neither firm can substitute between capital and labor. However, to the extent that consumers perceive the two firms as substitutes, the aggregate labor share decreases when the technology  $A_k$  of the capital-intensive firm increases, because this firm gains market share. Thus, the economy aggregates as if the elasticity of substitution in production is higher than one.

### Product Markets

Changes in product market structure are captured by the markup  $\mu$  in our labor share solution. An increase in the markup of price over marginal cost causes a decline of the labor share. The economics are that firms exploit an increase in product market power by increasing their prices and lowering their demand for inputs, thus leaving labor with a lower share of total income.

Before discussing the evidence for changes in product market markups, it is useful to clarify some terminology on income shares in relation to the measurements discussed earlier. In a perfectly competitive economy with constant returns to scale, labor and capital exhaust all income. With product and labor market imperfections or with decreasing returns to scale, firms make economic profits. In these cases, income is split into three components: the labor share, the capital share, and the profit share. What I called “compensation to capital” in the earlier discussion consists of both opportunity costs of using capital in production and economic profits. To be clear, the line labeled “Corporate Profits” in Table 1 is accounting profits and should not be confused with economic profits, a term that refers to revenues that exceed the sum of explicit and implicit costs of production.

What evidence do we have that markups have increased? De Loecker, Eeckhout, and Unger (2020) observe that the cost of goods sold has fallen dramatically relative to revenues during the past 40 years, and thus conclude that markups increased from roughly 20 percent to 60 percent. The increase in markups is driven by the small subset of firms at the top of the markup distribution and by a reallocation of economic activity toward high markup firms. How compelling is their evidence? My personal view is that product market markups have indeed increased over time, and this increase has had important consequences for the labor share and other macroeconomic outcomes. However, the magnitude of rising markups in this particular study seems overstated.<sup>8</sup>

<sup>8</sup>For a critical evaluation of the evidence on product-market markups in this journal, see Basu (2019). Online Appendix D in this paper summarizes my critique of De Loecker, Eeckhout, and Unger (2020). For example, the Compustat data used in their study cover only about 30 percent of economic activity.

Other evidence for increasing aggregate markups comes from the observation that aggregate investment rates and capital-to-output ratios are stable despite a decline of the labor share. If the decline of the labor share is not offset by an equal increase in the capital share, then economic profits must have risen.<sup>9</sup> Conversely, if increasing markups are the sole determinant of the labor share decline, then we would observe a decline of the capital share that is proportional to the decline of the labor share. Karabarbounis and Neiman (2014b) use this logic to attribute around half of the decline of the labor share to increasing markups, with the other half attributed to the decline of the relative price of capital goods. They calculate that markups increased by roughly 6 percentage points at the global level during the last four decades. Barkai (2020) also used estimates of the capital share to infer markups, and found that markups increased by roughly 17 percentage points in the United States. While these methods appear intuitive, calculating capital shares is complicated by factors such as unmeasured investments, adjustment costs to changing capital, and risk in capital accumulation (Karabarbounis and Neiman 2019).

Another piece of evidence concerns the role of “superstar” firms (Autor et al. 2020; Kehrig and Vincent 2021). The evidence shows that an increasing share of economic activity has been concentrated at larger firms with lower labor shares. This compositional effect generates a decline of the aggregate labor share. While it is tempting to do so, we should not use this evidence to conclude that product markets are becoming more imperfect over time. Industry concentration is an outcome and not a primitive cause. In fact, in the simple example of the two firms that I sketched before, markets are perfectly competitive, yet a technological change that favors larger and capital-intensive firms increases product market concentration and reduces the labor share. The industrial organization literature has identified several channels by which concentration and market power may be either positively or negatively correlated (Syverson 2019).

### **Labor Markets**

Changes in labor market structure are captured by the markup  $\theta$  in our labor share solution. An increase in the labor market markup causes a decline of the labor share. The economics are that firms exploit an increase in their labor market power by lowering their wages and demand for labor, thus leaving labor with a lower share of total income.

How should we think about the labor market markup? I note that the framework applies to any distortion on the labor side as long it augments the perceived

---

Also, those data include only firm sales, not prices and quantities of goods, and do not allow for a clean separation of fixed from variable costs.

<sup>9</sup>We should not necessarily equate economic profits to markups, because economic profits also increase when returns to scale decrease. A stable profit share is consistent with increasing markups if returns to scale also increase (for example, because of increasing fixed costs). As I show in online Appendix C, the distinction between profits and markups is important for evaluating the dynamic efficiency of an economy’s capital accumulation.



cost of labor, which equals  $(1 + \theta)w\ell$ , but it does not accrue to labor in the form of compensation, which is  $w\ell$ . Examples of labor distortions that affect the division of income are monopsony in which firms internalize workers' labor supply, oligopsony in which firms strategically interact in local labor markets, and wage bargaining between firms and workers. Some of these factors may also indirectly affect the labor share by limiting technology adoption and automation. Payroll taxes are an example of a distortion that does not affect the measured division of income, because they are included in the measure of compensation to employees. Another example is the monopoly union model, in which firms are price takers in labor markets but workers charge a markup that increases the wage above the opportunity cost of working. These distortions or rents are isomorphic to a change in the preference of workers to supply labor, in the sense that they affect the levels of compensation to employees and output, but not the ratio of compensation to employees to output.

What evidence do we have that labor market markups have contributed to the decline of the labor share? One good example is the New Deal policies during the Great Depression, which increased the bargaining power of workers and, as some series in Figure 1 show, coincided with an increase in the labor share. A second example is the Hartz reforms in Germany during the early 2000s, which reduced the bargaining power of workers and accelerated the decline of the labor share before the Great Recession, as Figure 2 shows. A third example is the worldwide decline of the share of employees with the right to bargain, which has declined by roughly 7 percentage points for the average country since the early 2000s (according to OECD data). While it seems plausible to attribute the decline of the labor share to changes in collective bargaining institutions, the evidence is inconclusive. The labor share has also declined in regions with high and relatively stable share of employees with the right to bargain, such as Austria, Belgium, countries in southern Europe, and countries in Scandinavia. For some other evidence on worker's bargaining power and the labor share, see Blanchard and Giavazzi (2003) and Elsby, Hobijn, and Sahin (2013).

In terms of labor market policies, it might be tempting to attribute the decline of the labor share to declines of the minimum wage, because a higher minimum wage removes labor market power from firms if markets are monopsonistic. A first read of the evidence on this point is also inconclusive. According to OECD data, in the past two decades, Australia, France, and the United States have experienced declines of their minimum wage relative to their median or average wage. However, Canada, Korea, and Spain have experienced increases. Yet, as shown in Figure 2, all six countries experienced declines of the labor share, although to varying degrees and at different times.

Recent work uses administrative datasets at the establishment level to infer labor market power and concludes that labor market power is decreasing over time. For examples, see Berger, Herkenhoff, and Mongey (2022) for the US nonfarm business sector, Yeh, Macaluso, and Hershbein (2022) for the US manufacturing sector, and Brooks et al. (2021) for the manufacturing sector in China and India. This evidence

is more compelling in my opinion, as one would reasonably expect labor to become increasingly mobile over time for various reasons. Increased labor mobility can be attributed to technological changes such as the internet, which decreased the cost of searching for jobs, to socioeconomic changes such as the decline of marriage and fertility rates, or, more recently, to work from home arrangements.

### **Globalization**

Globalization refers to the increase in the flow of goods, services, and financial assets across national borders during the last half-century or so. It is a multidimensional process and thus might potentially affect the labor share in a number of ways.

Globalization, outsourcing, and financial integration cannot have contributed to the decline of the labor share through changes in the distribution of economic activity across countries. After all, many different countries, ranging from advanced economies to emerging markets, and many different industries, ranging from manufacturing to nontraded sectors, have experienced declines of their labor share. If we think of emerging markets as relatively labor-abundant economies opening up to trade, one might predict a decline of the labor share in advanced economies due to a reallocation of economic activity away from labor-intensive sectors in advanced economies. However, this mechanism would counterfactually predict an increase in the labor share of emerging markets. Similarly, as the world removes restrictions on capital mobility, we expect the cost of capital to converge across countries. While this mechanism rationalizes the decrease in the labor share of emerging markets, it also counterfactually predicts an increase in the labor share of advanced economies.

Globalization might affect the labor share in other ways. In capital markets, globalization lowers transaction costs on financial assets and trade costs on imported capital goods, contributing to declines of the cost of capital for all countries. In production, globalization lowers trade costs of imported intermediate inputs for all countries. Standard economic models predict that lower prices of imported intermediate inputs cause a decline of the labor share if labor is more substitutable to these inputs than capital is (as discussed in online Appendix C). But many imported goods are capital goods, so it is not obvious if labor is more substitutable than domestic capital goods to imported capital goods. Globalization affects the structure of product markets, because it allows consumers to access a larger variety of imported goods and it pressures less productive domestic firms to exit. Globalization also affects technology adoption through foreign direct investments.

Overall, my view is that the role of globalization with regard to the falling labor share remains an open question, as we do not yet have a framework that allows us to discriminate between all these possibilities while accounting for the observed patterns of the labor share across different countries and industries.

## Implications of the Labor Share Decline

While some explanations, such as technological change and changes in product markets, appear more promising than others, it is fair to say that there is no strong consensus yet about the deeper causes of the labor share decline. With this caveat in mind, I organize the discussion of the implications of the labor share decline in three themes: what we learn about economic models, what we learn about aggregate welfare, and what the distributional consequences are.

### Economic Models

For a long time, economists have resisted giving up on models that generate constant macroeconomic ratios along the “balanced growth” path. Balanced growth is a property of a dynamic economic model in which variables eventually grow at a constant rate. For example, in a standard neoclassical growth model with exogenous growth of technology and population, output, capital, consumption, and investment grow at the exogenous growth rate of technology and population, labor grows at the population growth rate, and hours per worker are constant. Thus, the labor share is constant.

The decline of the labor share poses a challenge to models designed to be consistent with constant macroeconomic ratios along the balanced growth path. There are three ways to proceed. First, in certain settings it may be reasonable to ignore the problem, if the structural factors causing the decline of the labor share have nothing to do with the research question. This option is attractive, because models with nonbalanced growth are complicated and sometimes nonintuitive. Second, a researcher might maintain key assumptions that allow the model to be consistent with constant ratios and balanced growth in the long run, but allow the economy to deviate from this path at times because of transitory changes. The third option is to stop resisting and give up the assumptions that generate constant ratios and balanced growth in the long run. For example, a researcher could relax the assumption of an aggregate production function with unitary elasticity of substitution between capital and labor and allow for continuing capital-augmenting or investment-specific technological progress that leads the labor share to approach either zero or one in the long run.<sup>10</sup>

The findings that I discussed about product and labor market power have spurred further interest in developing economic and quantitative tools for studying models with imperfect competition. A key limitation of many of these models is that their predictions are sensitive to arbitrary definitions of what constitutes a market. For example, most economists would agree that not all manufacturing firms belong to

<sup>10</sup>A parallel is with the labor supply literature, in which economists postulate preferences so that income and substitution effects of wage changes cancel out in the long run. Thus, hours per worker remain constant despite the continuing improvement of technology. Boppart and Krusell (2020) develop a preference specification so that hours decline at a constant rate over time, as roughly observed in several countries.

one product market and that not all college-educated workers compete in one labor market. Conceptually, one could argue that every product or worker in the world constitutes a separate market because it is not strictly identical to any other product or worker. But this is not a useful disaggregation of economic activity. So, where do we draw the line? One promising alternative is to think about competition in the space of characteristics that define products, as in Pellegrino (2023).

Finally, models in corporate finance are often based on the idea that firms need to access external capital markets for financing. Such models may postulate some form of financial friction that restricts firms from accessing capital. But the decline of the labor share has been associated with changes in the distribution of available resources between households and corporations. Companies as a group have not distributed a significant fraction of their additional resources as payouts to households, but instead have retained them in the corporate sector in the form of holding cash, accumulating other assets, and paying back debt (Chen, Karabarbounis, and Neiman 2017). In fact, retained earnings have increased to such an extent that the corporate sector as a whole can finance its entire investments using retained earnings. Retention of earnings is likely to be heterogeneous across firms and countries, but the fact that many large companies need not rely on external capital markets calls for some rethinking of models that rely on frictions to generate interesting corporate financing and investment decisions.

### **Aggregate Welfare and Policy Implications**

The welfare implications of the labor share decline depend on which factors are primarily contributing to its decline. Suppose, first, that the decline reflects technological factors that allow economies to substitute away from labor toward tangible capital, intangible capital, and automation. Because economies choose a different mix of inputs when the old mix of inputs is available, aggregate welfare must increase. In this case, the decline of the labor share is a symptom of a more efficient aggregate technology of production.<sup>11</sup>

Alternatively, suppose that the decline of the labor share partly reflects changes in product or labor markets. It is fair to say that we do not yet have a complete understanding of changes in product and labor markets. Are barriers to entering product markets increasing over time? Are barriers to accessing particular occupations increasing over time? Are markups increasing because consumers are becoming less price sensitive, or are deregulation and merger policies to blame? To the extent that lower price sensitivity and deregulation allow firms to increase their margins without fostering innovation, one would expect aggregate welfare losses. There are several sources of efficiency losses arising from product and labor market markups. The textbook example of inefficiency is the deadweight loss that

<sup>11</sup> A caveat is that the lower cost of capital may have induced economies to accumulate more capital than is socially efficient. Following Abel et al. (1989), my preliminary tests show that this is not the case for many countries, as measured investment rates have not generally exceeded measured capital shares (see online Appendix B).

arises from distorting the marginal revenue product of factors relative to their price. Another example is the inefficiency in production if firms with market power have higher costs than if they operated in competitive markets. A final source of inefficiency arises from the misallocation of inputs when firms have different markups. Useful starting points are Atkeson and Burstein (2008), Peters (2020), and Edmond, Midrigan, and Xu (2023).

If the decline of the labor share is a symptom of an efficient change in aggregate technology, then there is nothing to do from the point of view of a policymaker who is interested in efficiency. However, if the labor share decline reflects changes in market power over time, policymakers should be concerned about the ability of markets to achieve efficient production and input levels. Policies that remove barriers to entry, foster competition, and subsidize production, potentially as a function of the scale of production, may help restore efficiency.

However, the welfare implications of the labor-share decline are subtle if technological progress is changing both the nature of production and the structure of product markets. For example, increased fixed costs in intangibles necessitate higher markups for firms to recoup their investments. If intangibles spill over across firms, then markups act as a subsidy to innovation. It is difficult to discriminate between the roles of technology and product market structure for the decline of the labor share, because technological changes affect product markups, product market structure may itself affect the direction of technical change, and policies and demographic trends may affect both of them at the same time. Further, one could be conflating the roles of technology and product market structure because some intangible investments are not measured in GDP but affect corporate valuations and welfare. Research aiming to separate these factors is extremely valuable.<sup>12</sup>

### **Distributional Effects and Policy Implications**

The decline of the labor share has important distributional consequences. Irrespective of what is causing the labor share to decline, a larger portion of income is absorbed by the capital factor that is more dispersed across households. We thus expect the decline of the labor share to be associated with more income inequality. Even if we accept that technological progress has increased standards of living on average, we would need to recognize that the benefits of technological progress are distributed unevenly across households. For example, the reduction in the price of capital goods increases the labor income gap between skilled and unskilled labor, because skilled labor is more complementary to capital than unskilled labor is (Krusell et al. 2000). Advances in information and communication technologies

<sup>12</sup>An example is Aghion et al. (2023), who show how an improvement of best firms' productivity is associated with lower long-run innovation and welfare losses. Other examples include Hopenhayn, Neira, and Singhania (2022), who argue that the welfare effects are not obvious if the decline in labor share reflects a decline of population growth, and Covarrubias, Gutiérrez, and Philippon (2020), who, in light of stable investment rates, interpret increasing concentration as reducing welfare. Crouzet and Eberly (2023) argue that opportunity costs and rents that accrue to unmeasured intangible investments help to reconcile stable physical investment rates with increasing corporate valuations.

have reduced the share of income accruing to routine labor (Eden and Gaggl 2018). In addition to disproportionately benefiting high-skilled labor, automation increases the return to wealth for capital owners (Moll, Rachel, and Restrepo 2022).

From a redistributive perspective, the policy implications of the labor share decline depend on whether it is transitory or will continue in the future. If it is transitory, insurance mechanisms that operate, explicitly, through asset markets and the usual government tax and transfer policies or, implicitly, through intrafamily transfers and substitution across firms, occupations, and industries will carry most of the adjustment. Acemoglu and Restrepo (2018) show that, if labor has a comparative advantage in creating new and more complex ideas, then the labor share need not decrease forever even if older tasks performed by labor are continuously automated. They use an interesting example from the technologies of the early twentieth century that made horses redundant after the introduction of machines: unlike horses, humans will not become redundant to machines because they have a comparative advantage in creating more complex tasks.

I am not as optimistic as Acemoglu and Restrepo (2018). Brynjolfsson and McAfee (2014) argue that the second machine age involves the automation of cognitive tasks that make humans and software-driven machines substitutes, rather than complements. Artificial intelligence may soon have a comparative advantage in producing new ideas and replace human labor even in complex tasks, as machines replaced horses a century ago. If technological advancements continue to favor capital indefinitely, the natural outcome is a transition to a world in which capital on its own produces the entire global income. Perhaps then, a natural policy response will be to institutionalize a national dividend, or even a global dividend, which guarantees a reasonable minimum standard of living to every person in the world. I call this policy a dividend as opposed to universal basic income for two reasons. The first reason is semantics, as output is entirely produced by capital. The second is more substantial. Universal basic income is financed by taxes which distort incentives to produce. However, one could hypothesize that artificial intelligence does not suffer disutility from producing and that humans maintain command of output. In that case, complete redistribution is feasible as there is no distortion from taxing production.

■ *The views in this paper do not necessarily reflect the views of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.*

## References

- Abel, Andrew B., Gregory Mankiw, Lawrence H. Summers, and Richard Zeckhauser. 1989. "Assessing Dynamic Efficiency: Theory and Evidence." *Review of Economic Studies* 56 (1): 1–19.

- Acemoglu, Daron, and David Autor.** 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings," In *Handbook of Labor Economics*, Vol. 4B, edited by David Card and Orley Ashenfelter, 1043–1171. Amsterdam: North-Holland.
- Acemoglu, Daron, and Pascual Restrepo.** 2018. "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review* 108 (6): 1488–1542.
- Acemoglu, Daron, and Pascual Restrepo.** 2022. "Tasks, Automation, and the Rise in U.S. Wage Inequality," *Econometrica* 90 (5): 1973–2016.
- Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li.** 2023. "A Theory of Falling Growth and Rising Rents." *Review of Economic Studies* 90 (6): 2675–2702.
- Atkeson, Andrew.** 2020. "Alternative Facts Regarding the Labor Share." *Review of Economic Dynamics* 37 (S1): S167–80.
- Atkeson, Andrew, and Ariel Burstein.** 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review* 98 (5): 1998–2031.
- Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen.** 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics* 135 (2): 645–709.
- Barkai, Simcha.** 2020. "Declining Labor and Capital Shares." *Journal of Finance* 75 (5): 2421–63.
- Barro, Robert J.** 2021. "Double Counting of Investment." *Economic Journal* 131 (638): 2333–56.
- Basu, Susanto.** 2019. "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence." *Journal of Economic Perspectives* 33 (3): 3–22.
- Berger, David, Kyle Herkenhoff, and Simon Mongey.** 2022. "Labor Market Power." *American Economic Review* 112 (4): 1147–93.
- Bernanke, Ben.** 2005. "The Global Saving Glut and the U.S. Current Account Deficit," Speech, Homer Jones Lecture, St. Louis, Missouri, April 14, 2005.
- Blanchard, Olivier, and Francesco Giavazzi.** 2003. "Macroeconomic Effects of Regulation and Deregulation in Goods and Labor Markets." *Quarterly Journal of Economics* 118 (3): 879–907.
- Boppart, Timo, and Per Krusell.** 2020. "Labor Supply in the Past, Present, and Future: A Balanced Growth Perspective." *Journal of Political Economy* 128 (1): 118–57.
- Brooks, Wyatt J., Joseph P. Kaboski, Yao Amber Li, and Wei Qian.** 2021. "Exploitation of Labor? Classical Monopsony Power and the Labor's Share." *Journal of Development Economics* 150: 102627.
- Brynjolfsson, Erik, and Andrew McAfee.** 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton.
- Chen, Peter, Loukas Karabarbounis, and Brent Neiman.** 2017. "The Global Rise of Corporate Saving." *Journal of Monetary Economics* 89: 1–19.
- Chirinko, Robert S.** 2008. " $\sigma$ : The Long and Short of It." *Journal of Macroeconomics* 30 (2): 671–86.
- Cooley, Thomas F., and Edward C. Prescott.** 1995. "Economic Growth and Business Cycles." In *Frontiers of Business Cycle Research*, edited by Thomas F. Cooley, 1–38. Princeton: Princeton University Press.
- Covarrubias, Matías, Germán Gutiérrez, and Thomas Philippon.** 2020. "From Good to Bad Concentration? US Industries over the Past 30 Years." In *NBER Macroeconomics Annual*, Vol. 34, edited by Martin Eichenbaum, Erik Hurst, and Valerie Ramey, 1–46. Chicago: University of Chicago Press.
- Crouzet, Nicolas, and Janice Eberly.** 2023. "Rents and Intangible Capital: A Q+ Framework." *Journal of Finance* 78 (4): 1873–1916.
- De Loecker, Jan, Jan Eeckhout, and Gabriel Unger.** 2020. "The Rise of Market Power and Macroeconomic Implications." *Quarterly Journal of Economics* 135 (2): 561–644.
- Eden, Maya, and Paul Gaggl.** 2018. "On the Welfare Implications of Automation." *Review of Economic Dynamics* 29: 15–43.
- Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu.** 2023. "How Costly Are Markups?" *Journal of Political Economy* 131 (7): 1619–75.
- Eggertsson, Gauti B., Neil R. Mehrotra, and Jacob A. Robbins.** 2019. "A Model of Secular Stagnation: Theory and Quantitative Evaluation." *American Economic Journal: Macroeconomics* 11 (1): 1–48.
- Eisfeldt, Andrea L., Antonio Falato, and Mindy Z. Xiaolan.** 2023. "Human Capitalists." In *NBER Macroeconomics Annual*, Vol. 37, edited by Martin Eichenbaum, Erik Hurst, and Valerie Ramey, 1–61. Chicago: University of Chicago.
- Elsby, Michael W. L., Bart Hobijn, and Ayşegül Sahin.** 2013. "The Decline of the U.S. Labor Share." *Brookings Papers on Economic Activity* 44 (2): 1–52.

- Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer.** 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105 (10): 3150–82.
- Fleck, Susan, Steven Rosenthal, Matthew Russell, Erich H. Strassner, and Lisa Usher.** 2014. "A Prototype BEA/BLS Industry-Level Production Account for the United States," In *Measuring Economic Sustainability and Progress*, edited by Dale W. Jorgenson, J. Steven Landefeld, and Paul Schreyer, 323–72. Chicago: University of Chicago Press.
- Gollin, Douglas.** 2002. "Getting Income Shares Right." *Journal of Political Economy* 110 (2): 458–74.
- Gomme, Paul, B. Ravikumar, and Peter Rupert.** 2011. "The Return to Capital and the Business Cycle." *Review of Economic Dynamics* 14 (2): 262–78.
- Greenwood, Jeremy, Zvi Hercowitz, and Per Krusell.** 1997. "Long-Run Implications of Investment-Specific Technological Change." *American Economic Review* 87 (3): 342–62.
- Güvenen, Fatih, Raymond J. Mataloni Jr., Dylan G. Rassier, and Kim J. Ruhl.** 2022. "Offshore Profit Shifting and Aggregate Measurement: Balance of Payments, Foreign Investment, Productivity, and the Labor Share." *American Economic Review* 112 (6): 1848–84.
- Hall, Robert E., and Dale W. Jorgenson.** 1967. "Tax Policy and Investment Behavior." *American Economic Review* 57 (3): 391–414.
- Henry, Peter Blair.** 2003. "Capital-Account Liberalization, the Cost of Capital, and Economic Growth." *American Economic Review* 93 (2): 91–96.
- Hopenhayn, Hugo, Julian Neira, and Rish Singhania.** 2022. "From Population Growth to Firm Demographics: Implications for Concentration, Entrepreneurship and the Labor Share." *Econometrica* 90 (4): 1879–1914.
- Hubmer, Joachim.** 2023. "The Race between Preferences and Technology." *Econometrica* 91 (1): 227–61.
- Hurst, Erik, and Benjamin Wild Pugsley.** 2011. "What Do Small Businesses Do?" *Brookings Papers on Economic Activity* 42 (2): 73–118.
- Kaldor, Nicholas.** 1957. "A Model of Economic Growth." *Economic Journal* 67 (268): 591–624.
- Karabarbounis, Loukas.** 2024. "Replication data for: Perspectives on the Labor Share." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E198526V1>.
- Karabarbounis, Loukas, and Brent Neiman.** 2014a. "Capital Depreciation and Labor Shares around the World: Measurement and Implications." NBER Working Paper 20606.
- Karabarbounis, Loukas, and Brent Neiman.** 2014b. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129 (1): 61–103.
- Karabarbounis, Loukas, and Brent Neiman.** 2019. "Accounting for Factorless Income." In *NBER Macroeconomics Annual*, Vol. 33, edited by Martin Eichenbaum, Erik Hurst, and Valerie Ramey, 167–228. Chicago: University of Chicago Press.
- Kaymak, Barış, and Immo Schott.** 2023. "Corporate Tax Cuts and the Decline in the Manufacturing Labor Share." *Econometrica* 91 (6): 2371–2408.
- Kehrig, Matthias, and Nicolas Vincent.** 2021. "The Micro-Level Anatomy of the Labor Share Decline." *Quarterly Journal of Economics* 136 (2): 1031–87.
- Keynes, J. M.** 1939. "Relative Movements of Real Wages and Output." *Economic Journal* 49 (193): 34–51.
- Koh, Dongya, Raúl Santaaulàlia-Llopis, and Yu Zheng.** 2020. "Labor Share Decline and Intellectual Property Products Capital." *Econometrica* 88 (6): 2609–28.
- Krusell, Per, Lee E. Ohanian, José-Victor Ríos-Rull, and Giovanni L. Violante.** 2000. "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis." *Econometrica* 68 (5): 1029–53.
- Mian, Atif, Ludwig Straub, and Amir Sufi.** 2021. "Indebted Demand." *Quarterly Journal of Economics* 136 (4): 2243–2307.
- Moll, Benjamin, Lukasz Rachel, and Pascual Restrepo.** 2022. "Uneven Growth: Automation's Impact on Income and Wealth Inequality." *Econometrica* 90 (6): 2645–83.
- Pellegrino, Bruno.** 2023. "Product Differentiation and Oligopoly: A Network Approach." <http://dx.doi.org/10.2139/ssrn.3329688>.
- Peters, Michael.** 2020. "Heterogeneous Markups, Growth, and Endogenous Misallocation." *Econometrica* 88 (5): 2037–73.
- Piketty, Thomas.** 2014. *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press.
- Ricardo, David.** 1817. *On the Principles of Political Economy and Taxation*. London: John Murray.
- Robinson, Joan.** 1933. *The Economics of Imperfect Competition*. New York: MacMillan.
- Rognlie, Matthew.** 2015. "Deciphering the Fall and Rise in the Net Capital Share: Accumulation or Scarcity?" *Brookings Papers on Economic Activity* 46 (1): 1–54.



- Solow, Robert M.** 1958. "A Skeptical Note on the Constancy of Relative Shares." *American Economic Review* 48 (4): 618–31.
- Syverson, Chad.** 2019. "Macroeconomics and Market Power: Context, Implications, and Open Questions." *Journal of Economic Perspectives* 33 (3): 23–43.
- Yeh, Chen, Claudia Macaluso, and Brad Hershbein.** 2022. "Monopsony in the US Labor Market." *American Economic Review* 112 (7): 2099–2138.



# Why Labor Supply Matters for Macroeconomics

Richard Rogerson

Understanding the forces that determine the level of GDP is one of the core objectives of macroeconomics, and the concept of the aggregate production lies at the center of this effort. The aggregate production function, typically written as  $Y = AF(K, L)$ , describes how inputs—labor ( $L$ ) and capital ( $K$ )—are used to produce GDP ( $Y$ ) given the existing level of technology ( $A$ ). The desire to understand the forces that determine the level of GDP naturally leads one to consider the forces that shape the level of inputs. For this reason, labor features prominently in all undergraduate textbooks in macroeconomics.

Given the importance of labor as a determinant of GDP, a student who has taken a course in microeconomics might reasonably expect that a standard demand and supply model of the labor market would be an integral part of a larger macroeconomic model. But this student will be surprised to find that the concept of labor supply is completely absent from the two key models that most undergraduate macro textbooks teach to students: a version of the Solow-style growth model and the IS–LM model. The Solow model is introduced as the dominant framework for thinking about the long run level of GDP in the economy. This model has no discussion of the labor market, and asserts that the quantity of labor is simply proportional to the size of the population. The IS–LM model is introduced as the dominant

■ *Richard Rogerson is the Charles and Marie Robertson Professor of Public and International Affairs, Princeton School of Public and International Affairs, Princeton, New Jersey. His email address is [rdr@princeton.edu](mailto:rdr@princeton.edu).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.2.137>.

framework for understanding business cycle fluctuations in labor, but it also has no discussion of labor supply. It implicitly assumes that the quantity of labor in the market is completely determined by demand.

In this paper, I argue that labor supply indeed matters for macroeconomics, and that its omission from undergraduate textbooks is an important one. I make this argument in the context of understanding long-run differences in hours of work across advanced economies. In the first part of this paper, I document three facts. First, hours of work vary substantially across countries, even after controlling for differences in population, and the magnitude of these differences imply large effects on GDP. Second, incentives to supply labor across country also vary substantially, as proxied with the magnitude of tax and spending programs. Third, there is a strong positive correlation between the level of hours of work per person and incentives to supply labor. These three facts constitute strong suggestive evidence that labor supply matters for macroeconomic outcomes.

So why do most undergraduate macro textbooks not incorporate labor supply as an important element? I think the answer lies in the prevailing wisdom about the magnitude of the *aggregate elasticity of labor supply*, which measures the magnitude of the change in aggregate labor supply in an economy in response to a change in incentives to work. If this elasticity is small, then labor supply responses might be small even in the presence of large changes in incentives. In a survey, Saez, Slemrod, and Giertz (2012) write: “[T]he profession has settled on a value for this elasticity close to zero for prime-age males. . . . [O]verall the compensated elasticity of labor seems to be fairly small.” Based on this prevailing wisdom, even large differences in work incentives have very little effect on labor supply, implying that labor supply considerations are not of first-order importance and can be safely ignored.

In the second part of this paper, I argue that this prevailing wisdom about the magnitude of the aggregate labor supply elasticity is based on a poor understanding of the underlying economics of aggregate labor supply. The key issue that I stress is the need to consider labor supply responses along both the intensive (hours per worker) and extensive (fraction of population employed) margins. Early estimates of labor supply elasticities focused only on the intensive margin for a particular demographic group. Relative to this baseline I make two points. First, as an empirical matter I show that the extensive margin is of greater importance. Second, as a theoretical matter I argue that one cannot assume that the labor supply elasticity for one demographic group applies to all demographic group. Taken together, these two points imply that a small elasticity for one demographic group along the intensive margin does not provide reliable information about the aggregate elasticity for the overall population.

### **Three Facts**

In this section, I document the three facts described in the introduction for a sample of 22 advanced economies: Australia, Austria, Belgium, Canada, Denmark,

Finland, France, Germany, Greece, Ireland, Italy, Japan, Korea, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom and the United States. All of the statistics reported below come from the OECD and represent averages for the period 2015–2019. I average over five years because I am interested in long-run differences, and this reduces the potential for the values to reflect differences in short-run outcomes due to business cycle fluctuations. I choose the period 2015–2019 to avoid the influence of the two most recent business cycle events—the Great Recession of 2008 and the pandemic recession of 2020, both of which had large effects in virtually all countries.

**Fact 1: Cross-Country Differences in Hours of Work Are Large**

Total hours of work are computed as the product of total employment and average annual hours of work per person in employment. The measure of average annual hours of work per person in employment reflects hours at work rather than hours paid for, and so is adjusted to exclude statutory holidays and vacation time.

I am particularly interested in exploring how total hours of work vary across countries after controlling for population differences. In what follows, I will use the size of the population that is 15 years and older as a measure of population, because it seems reasonable to exclude individuals younger than 15 when considering market work in this set of advanced economies.<sup>1</sup> I refer to the ratio of total hours of work to the size of the population 15 and older as *hours per person*. Importantly, this measure is distinct from *hours per worker*.

Hours per person varies from 758 hours per year in Italy to 1,230 hours per year in Korea. The United States has a value of 1,096 hours per person, which is third-highest among our sample of 22 countries. To facilitate comparison of these values across countries, it is useful to report values relative to the United States, and Table 1 reports these values.

Table 1 documents very large differences in hours of work per person across these economies. For example, hours per person in Italy are more than 30 percent less than in the United States. If one were to assume a commonly used back-of-the-envelope Cobb–Douglas production function  $Y = K^{(1/3)}L^{(2/3)}$ , this would amount to a GDP difference in excess of 20 percent, a difference that is almost five times as large as the drop in real GDP per capita experienced by the United States in the Great Recession! While the United States and Italy are two of the more extreme values, they are not outliers: many countries have values very similar to Italy and several are similar to the United States.

How should we think about these large differences? One might conjecture that they reflect differences in unemployment rates across countries. This turns out not to be the case. One way to explore this possibility is to do a counterfactual

<sup>1</sup>One might also consider excluding individuals older than some age. In many countries, individuals older than 65 are relatively rare in the workforce, but in several countries, many individuals work beyond this age. The main message from the results reported below would be unaffected if we were to instead use the size of the population aged 15–64 or 15–74.

Table 1

**Hours of Work per Person Relative to the United States**

Considerably below the US level (<0.75)	Moderately below the US level (0.75, 0.85)	Slightly below the US level (0.85, 0.95)	At or above the US level (>0.95)
Italy (0.69)	Finland (0.77)	UK (0.85)	Canada (0.96)
France (0.70)	Austria (0.79)	Sweden (0.90)	Australia (0.98)
Belgium (0.72)	Norway (0.80)	Ireland (0.91)	United States (1.00)
Greece (0.73)	Netherlands (0.82)	Japan (0.91)	New Zealand (1.07)
Denmark (0.74)	Portugal (0.85)	Switzerland (0.93)	Korea (1.12)
Germany (0.74)			
Spain (0.75)			

*Source:* Author's calculation using data from OECD (2024a, c).

*Note:* Details of the calculation are in the online Appendix. Table shows average for 2015–2019.

calculation for each country in which we assume each unemployed individual becomes employed, and works the same number of hours as the average employed individual in their respective country. I do not report the results of this calculation here, but it turns out that this has only a minor effect on the differences reported in Table 1.<sup>2</sup> Put somewhat differently, large differences in hours per person across countries are driven by differences in participation rates and hours per worker, not by differences in the unemployment rates.

Given that differences in unemployment do not account for the large differences in work across countries, a natural characterization of the situation is that people in different countries are making different choices about work–life balance, with countries in the right-most column of Table 1 choosing to have less leisure and more consumption relative to the countries in the left most column. Economics offers two broad explanations for why people might make different choices. One is that preferences about work–life balance are systematically different across countries. The other is that differences in policy across countries change the costs and benefits of choices regarding work–life balance in such a way that they lead individuals to make systematically difference choices across countries.

While each of these two possibilities ascribes an important role for labor supply in accounting for the differences reported in Table 1, the two possibilities have very different implications for the broader role of labor supply in macroeconomics. If the large differences essentially reflect responses of labor supply to differences in policy, then it follows that analyses of macroeconomic policy need to be mindful of potential responses in labor supply. Alternatively, if the large differences observed across countries essentially reflect permanent differences in preferences across countries, and macroeconomic policy does not affect preferences, perhaps labor supply issues can be left to the sociologists—and it is no longer clear that one needs to include

<sup>2</sup>Results are reported in the online Appendix, Table A1.

labor supply responses in an analysis of macroeconomic policy. To address this issue, we next turn to documenting differences in policy across countries.

**Fact 2: Cross-Country Differences in Work Incentives Are Large**

A very large literature has documented substantial differences in the scale and extent of many tax and transfer programs across countries, noting how these programs affect the incentives to supply labor. Prominent examples of such transfer programs include social security, disability insurance, unemployment insurance, and general social assistance. All of these programs involve cash transfers, but similar incentive effects are associated with government provision of goods and services, such as education and health care. Each relevant program is associated with a rich set of details regarding the taxes used to finance them, who is eligible, and the nature and extent of benefits that are paid. The specific details of each program can influence the extent to which they influence work incentives.

Describing the details of all of these programs is beyond the scope of this paper. However, because a more expansive tax and spending program requires larger government revenues, I use the ratio of general government revenues to GDP as a crude measure of the size of tax and spending programs. The implicit assumption in what follows is that that larger tax and transfer programs are associated with lower work incentives.<sup>3</sup>

Once again, I compute the average value of this ratio over the period 2015–2019 for each of the 22 countries in the sample. Table 2 provides information on the distribution of this measure.

The differences in this measure across countries are large: nine countries have values that are less than 40 percent, while six countries, all from Europe, have values that exceed 50 percent. Many European countries have values that are roughly 20 percentage points higher than the United States.

**Fact 3: Work Incentives and Hours of Work Are Strongly Correlated**

Having documented differences in both the amount of work per person across countries and a broad measure of work incentives across countries, we now examine how the two are related. A comparison of Tables 1 and 2 readily suggests a strong negative correlation between the two measures. In fact, the correlation is equal to  $-0.72$ . To better visualize the relationship, Figure 1 shows a scatter plot of these two variables.

Figure 1 also includes an estimated regression line that results from regressing the log of hours per person on a constant and government revenues as a percent of GDP. This regression yields a coefficient of  $-0.012$  on government revenues, which implies that a 1 percentage point increase in government revenues

<sup>3</sup>I am necessarily taking a very broad brush view, and a richer analysis would recognize many nuances. For example, as both Rogerson (2007) and Kleven (2014) have argued, an important distinguishing feature of the relatively large tax and transfer systems in Scandinavia is that tax revenue is partially used to increase work incentives by subsidizing things like child and elderly care.

Table 2

**Government Revenues as a Percentage of GDP**

Low (<40%)	Medium (40–50%)	High (>50%)
Ireland (26.1%)	Canada (41.2%)	Sweden (50.2%)
United States (32.5%)	Portugal (42.9%)	Belgium (50.9%)
Korea (33.2%)	Netherlands (43.6%)	Denmark (52.6%)
Switzerland (34.4%)	Germany (45.8%)	France (53.1%)
Australia (35.2%)	Italy (46.8%)	Finland (53.2%)
Japan (35.4%)	Austria (49.1%)	Norway (55.8%)
United Kingdom (38.4%)	Greece (49.5%)	
Spain (38.7%)		
New Zealand (39.6%)		

Source: OECD (2024b).

Note: Results are averages for 2015–2019.

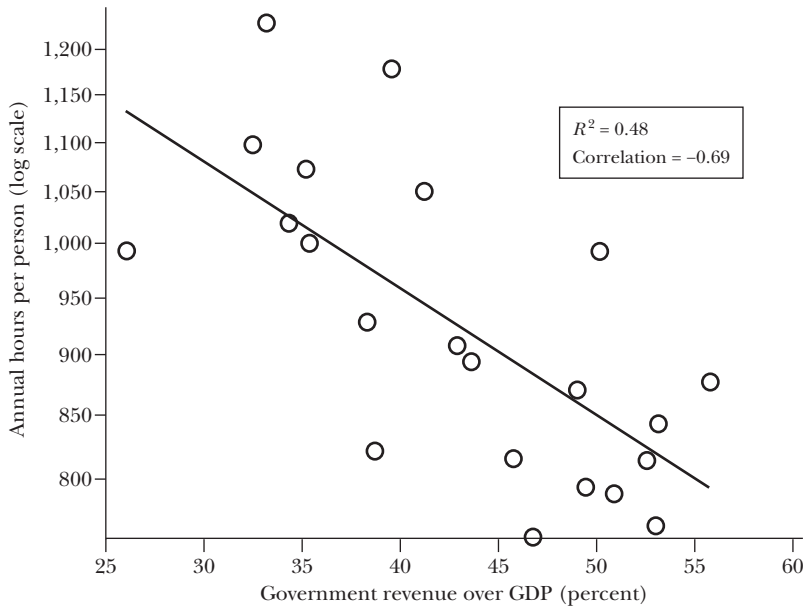
as percent of GDP is associated with a 1.2 percent decrease in hours per person. This estimate implies that a 20 percentage point difference in government revenues as a percentage of GDP is associated with a difference of 24 percent in hours per person. These magnitudes are close to the differences between the United States and several western European countries, like France, Denmark, and Belgium. This regression has an  $R^2$  of slightly more than 0.50, indicating that this broad measure of incentives has a substantial amount of explanatory power for hours per person.

Might preferences also display large systematic differences across these countries? After all, economists routinely assume that preferences vary across individuals as a way to explain why similar individuals make different choices despite facing the same options. In particular, this is how economists explain why individuals with similar demographic characteristics, wage rates, and wealth within a given country choose to work different hours. While many economists are sometimes reluctant to assume that preferences differ systematically across countries or other large groups of individuals, recent work has highlighted the potential for culture or social norms to generate systematically different choices across large groups of individuals (for example, Fernández and Fogli 2009).

Evidence regarding these effects typically examines how choices are persistently different across long periods of time; for example, examining the relationship between choices across generations. With this in mind, I briefly discuss one argument against the view that differences in preferences are the dominant source of the differences in Table 1. Key to this argument is that the cross-country differences documented in Table 1 were very different several decades ago. In particular, Ohanian, Raffo, and Rogerson (2008) document that in the 1960s, hours of work in many Western European countries were very similar to hours of work in the United States, and in some cases even greater. If differences in preferences are the dominant source of differences in hours of work across countries, this would require that preferences have changed systematically over time. Moreover, Ohanian, Raffo,



Figure 1

**Government Revenues and Hours of Work**

Source: Author's calculation using data from OECD (2024a, b, c).

Notes: Each point is a country. Values are averages for 2015–2019.

and Rogerson (2008) also show that the differential secular trends in hours of work across countries since the 1960s are highly correlated with the differential expansions of tax and spending systems. While this evidence does not rule out differences in preferences as a contributing factor, it does challenge the notion that differences in preferences are the dominant source of current differences.

If the level of work is an important input into production, one might be tempted to conclude that higher work incentives are an unambiguously good thing for an economy. However, answering this question calls for a more nuanced understanding of the role of work incentives. While lower work incentives can in some cases have a negative impact on overall economic welfare, there are also many examples of policies and programs that might improve economic welfare at the same time that they lower the incentive for some individuals to work.

A good case in point is programs that fall under the label of social insurance, like disability insurance or unemployment insurance. Consider the case of unemployment insurance. Just as fire and car insurance are valuable because they protect households against negative shocks, unemployment insurance improves economic welfare by protecting households against random loss of income due to factors that do not reflect on them. Laid-off workers who receive unemployment insurance will be able to take a greater period of time to find a new job that best matches their

abilities, whereas workers who do not receive unemployment insurance may need to take the first job that becomes available, even if this is not a good use of their abilities. Achieving a good match between job requirements and worker abilities is important for the efficient use of resources. In this sense, unemployment insurance may indeed lower the incentives for a laid-off worker to find work immediately, but may nonetheless promote economic efficiency.<sup>4</sup>

More generally, programs that foster redistribution may lower work incentives, but this may come at the benefit of generating a more equitable income distribution. Put somewhat differently, while lower work incentives may impose a cost on society, the programs that generate these lower incentives may also generate benefits, and so the evaluation of these programs must assess both the costs and the benefits. This paper focuses purely on the issue of whether changes in work incentives do indeed have significant effects on overall hours of work in the economy and should not be interpreted as an assessment of the overall desirability of programs that affect these incentives.

To summarize, the three facts presented in this section strongly suggest that labor supply responses to differing work incentives may be an important factor in understanding the large differences in hours of work across countries. But in completing this argument, one key piece of evidence is still missing: independent evidence on the magnitude of individual responses to the differences in work incentives associated with these different policies. The next section will discuss evidence on the magnitude of these labor supply elasticities.

## **The Aggregate Labor Supply Elasticity**

The cross-country evidence documenting a strong negative correlation between hours of work per person and the size of government spending could be viewed as indirect evidence about the magnitude of the aggregate labor supply elasticity. But economists prefer to find independent evidence regarding how individuals respond to changes in work incentives.

To determine the magnitude of the aggregate labor supply elasticity, one might ideally want to conduct a controlled experiment in which individuals face different work incentives and we observe how this affects their choices for hours of work. However, the differences that we report in Table 1 reflect differences in labor supply across people of all ages, and so reflect differences in lifetime choices that individuals have already made and are making. When considering situations that reflect lifetime choices, conducting such a controlled experiment in this area is nearly impossible.

Thus, economists have devised an alternative strategy. The idea is to combine data from the real world on the choices that individuals make over time, within the context of a model that specifies how individuals make their choices. We require

<sup>4</sup>See Acemoglu and Shimer (1999) for a rigorous analysis of this issue. In particular, they show that unemployment insurance can lead to higher output.

that the model of individual choice is consistent with the data that we observe. If we do this for a large sample of individuals, we can reliably use the model to predict the average individual response to specific changes in work incentives associated with a particular program. In particular, we could compute how lifetime hours will respond to a permanent change in work incentives. Such an exercise provides a source of independent evidence on the aggregate labor supply elasticity.

I emphasize that the model one uses will affect the inference of the researcher regarding the aggregate labor supply elasticity. In what follows, I describe two different models of labor supply and how they affect inference about the magnitude of the aggregate labor supply elasticity. But before presenting those models, it is useful to start with the canonical textbook model of labor supply.

### Canonical Textbook Model of Labor Supply

The canonical textbook model of individual labor supply considers an individual who makes a single continuous choice regarding hours of work subject to a linear budget equation. Specifically, the individual has utility  $U(c, h)$  that is defined over consumption ( $c$ ) and hours of work ( $h$ ). In what follows I will assume that the utility function takes the following form, which is commonly used in applied work:

$$U(c, h) = c^\sigma - \alpha h^\gamma.$$

This utility function features two key curvature parameters,  $\sigma$  and  $\gamma$ , that are relevant for labor supply elasticities. The parameter  $\sigma$  captures the declining marginal utility of consumption, while  $\gamma$  captures the declining marginal utility of leisure. The parameter  $\alpha$  is allowed to vary across individuals as a way to capture the possibility that individuals differ in their preferences regarding work–life balance. An individual with a high value of  $\alpha$  is someone with a relatively higher value on leisure relative to consumption.

The individual can choose any value for hours  $h$  in the interval  $[0, H]$ , where  $H$  is the total time endowment available, and total labor earnings will be the product of the wage rate  $w$  and their choice of hours  $h$ . I also specify a specific form for a stylized tax and transfer system. Individuals face a tax rate of  $\tau$  on labor earnings but also receive a transfer from the government equal to  $T$ , with the value of  $T$  dictated by a balanced budget constraint for the government, so that on average, all of the tax revenues are being transferred back to households, either as cash payments or as government-provided goods and services.<sup>5</sup>

If  $c$  is the quantity of consumption and  $p$  is the price per unit of consumption, then the individual's budget equation reads:

$$pc = (1 - \tau)wh + T.$$

<sup>5</sup>In this specification, the value of the transfer  $T$  is the same for all individuals. The *net transfer* that an individual receives from the government is equal to  $T - \tau wh$ , which is decreasing in income, so that this tax and transfer system is redistributing from higher income individuals to lower income individuals.

While very stylized, this representation of the tax and transfer system captures the fact that a higher value of  $\tau$  indicates a larger size of the tax and transfer system and serves to decrease work incentives. That is, in this setting, a higher value of  $\tau$  will lead an individual to choose a lower value for  $h$ . We are interested in the average magnitude of this response across all individuals.

In the canonical textbook model, the individual makes a single choice about hours of work. If we want to use this model to think about lifetime choices regarding work–life balance, we can interpret the choices in the model to reflect lifetime choices about consumption and hours of work. With this latter interpretation,  $h$  reflects hours of work over the individual’s lifetime, and the budget equation is implicitly imposing the constraint that spending on consumption over the lifetime cannot exceed lifetime income.

To understand the determinants of the aggregate labor supply elasticity, it will be useful to dig deeper into the details of lifetime labor supply. In particular, there are two different margins that people might adjust in altering the amount of time devoted to work over a lifetime: the *extensive* margin of how many years they spend in employment, and the *intensive* margin of how many hours they work in the years that they are employed. Understanding the role of these two margins will be very relevant for our discussion about the aggregate labor supply elasticity. As a starting point, I consider two different models of lifetime labor supply, one where only the intensive margin is active and the other where only the extensive margin is active, before considering a third model that features both margins.

### **An Intensive Margin Model**

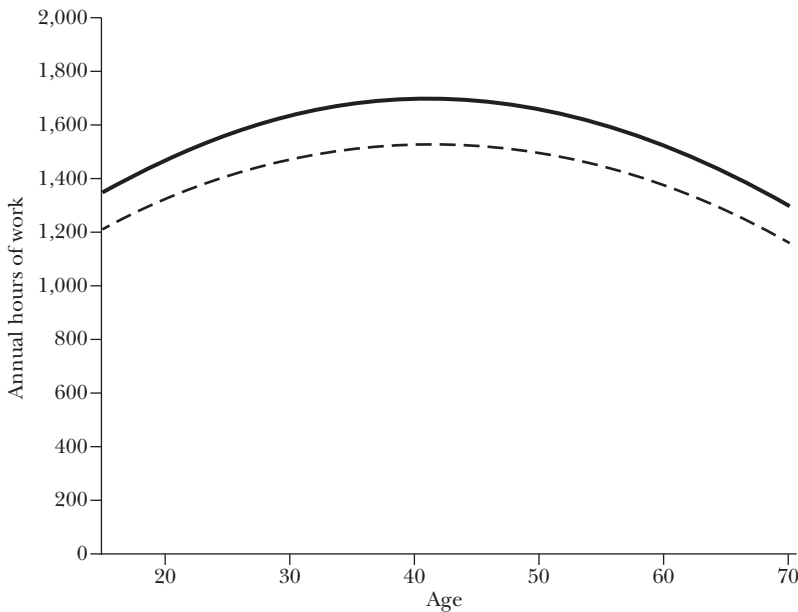
The intensive margin model of labor supply choices will look very much like the canonical textbook model, except that it is extended to allow for many periods. The individual’s lifetime labor supply problem becomes one of choosing how many hours to work in each period subject to a lifetime budget constraint.<sup>6</sup> For simplicity, assume that lifetime utility is just the sum of utility over all of the periods and that the interest rate is zero so that total lifetime spending and total lifetime income are also just the sum of spending and income over all of the periods.

If preferences, the wage rate, and the tax and transfer system are not changing over time, then this model predicts that the individual will choose to work the same number of hours in each period. If the tax and transfer system changes, the individual would change hours worked by the same amount in all periods. Importantly, differences in the tax and transfer system will influence how much the individual works on average but will not influence the number of years in which the individual is employed.

More generally, an individual’s wage, as well as the disutility that they associate with working, may change with age. In this case, holding the tax and transfer system fixed over time, the individual will have a lifetime profile for hours that also

<sup>6</sup>This model allows for the possibility that there is an exogenously specified time after which hours of work are always zero. The key property is that this time is held fixed.

Figure 2

**Lifetime Hours in the Pure Intensive Margin Model**

Source: Author's creation.

Note: Illustration of life cycle labor supply in intensive margin model. Dashed line shows the effect of an increase in the size of the tax and transfer system.

varies over time, and in particular will work more during periods when the wage is relatively high and/or the disutility of work is relatively low. However, a permanent change in the tax and transfer system will shift the entire profile up or down proportionately; that is, the percentage change in hours will be the same at all ages. This is illustrated in Figure 2. The dashed line indicates how the individual's choice of hours changes if the size of the tax and transfer system was increased. Importantly, in this model, all of the adjustment in lifetime hours comes from a change in the average hours per year while working and not from adjustment in the number of years worked.

**An Extensive Margin Model**

The extensive margin model looks quite different. In this economy, the individual does not freely choose how many hours to work in any given period. Instead, individuals can either choose to work a prespecified number of hours that we denote by  $\hat{h}$ , or they can choose not to work. One might think of  $\hat{h}$  as reflecting a 35- or 40-hour workweek. The key feature here is that the individual can choose whether to work, but cannot choose their hours conditional on deciding to work.

For this individual, the lifetime labor supply problem involves choosing how many years to spend in employment, subject to the lifetime budget constraint. While it is possible that an individual would choose to work every year throughout their lifetime, the empirically relevant case is one in which the level of  $\hat{h}$  is sufficiently high that individuals choose only to work for a fraction of the years. If the individual's wage and disutility of work vary with age, then the individual will choose to work in those periods in which the wage is relatively high and the disutility of work is relatively low. In this model, a change in the tax and transfer system will change the fraction of the individual's lifetime that they spend in employment, but by assumption will have no impact on how many hours the individual works when working. If wages and the disutility of work change relatively smoothly with age, then periods of work will tend to be bunched during one part of the life cycle.<sup>7</sup>

Figure 3 illustrates the previous discussion, with the dashed line again reflecting how the profile for lifetime hours adjusts in response to an increase in the size of the tax and transfer system.

Before discussing the properties of the dashed line, it is of interest to contrast the solid line in this extensive-margin diagram with the intensive-margin solid line in the previous diagram. These two curves are drawn so that total lifetime hours are approximately equal. The key difference is that lifetime hours of work are more concentrated in Figure 3 than in Figure 2. In the intensive margin model, the individual tends to smooth work across all available periods. But in the extensive margin model, the individual is not able to do this, because a worker cannot reduce hours below  $\hat{h}$  when employed.

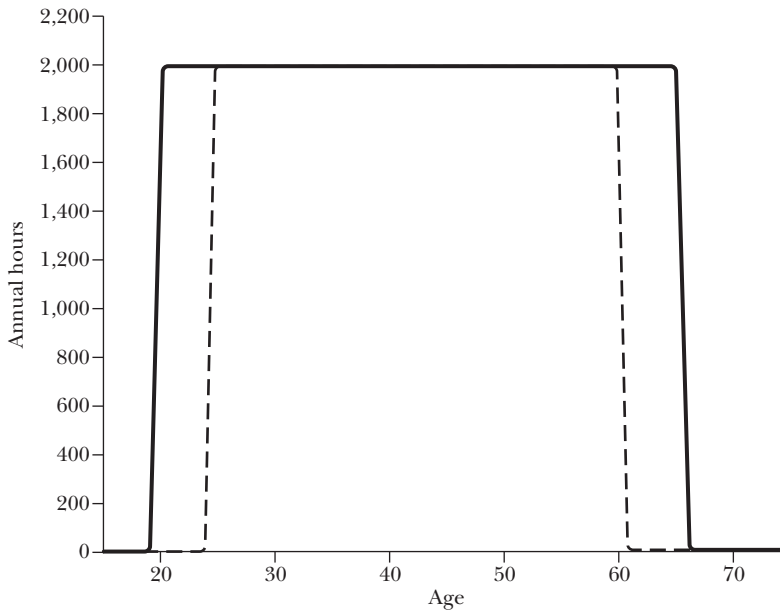
Turning now to consider the dashed line in Figure 3, we see that an increase in the scale of the tax and transfer program causes individuals to reduce the number of periods in which they work, but does not change the number of hours worked in periods in which they are employed. A key property is that the reduction in working time happens at the edges of the life cycle profile. To see why this is, note that the decision to work in a particular period entails a cost and a benefit. The cost is the disutility of working, and the benefit is the increase in income. A period in which the individual switches from working to not working is a point at which the net benefit to working moves from positive to negative. If the costs and benefits change continuously with age, then the last period of work before the individual switches to not working will be a period in which the net benefit is positive but very close to zero. The tax and transfer system reduces the benefit of working in each period but leaves the disutility of work unchanged. Thus, a period that previously had a small net benefit will now have a small net cost.

Importantly, Figure 3 simply illustrates the lifetime choices for a particular individual and corresponds to what one might think of as the traditional male lifetime working profile. Individuals with different preferences and different wages will potentially have different lifetime profiles for work. Many women experience

<sup>7</sup>In some situations, there will be large changes in one of these variables in a given period that could generate short periods of nonemployment. Childbirth would be an important example of this.

Figure 3

**Lifetime Hours in the Pure Extensive Margin Model**



Source: Author’s creation.

Note: Illustration of life cycle labor supply in extensive margin model. Dashed line shows the effect of an increase in the size of the tax and transfer system.

mid-career interruptions associated with childbirth, so their profile would look quite different. However, in an extensive-margin model, the general effect of an increase in the scale of the tax and transfer system is to expand the regions in which the individual does not work at all.

**Messages for Measuring the Aggregate Labor Supply Elasticity**

Figures 2 and 3 have implications for how to obtain measures of the aggregate labor supply elasticity, and as a result important implications for interpreting the vast majority of existing evidence. Starting in the 1980s, data became available that allowed researchers to follow the labor market choices of individuals over time in the United States and to use these data to estimate labor supply elasticities. The most impactful of these papers essentially adopted an intensive margin model as their framework.<sup>8</sup> Additionally, they focused on the behavior of annual hours of work for prime-aged males that worked in each year. The results of these estimation exercises

<sup>8</sup>The key data source in the United States was the Panel Study of Income Dynamics, a dataset that interviewed a given set of individuals each year starting in 1968, thereby creating a panel dataset on wages and

implied very small responses to a permanent change in work incentives. Importantly, in the intensive margin model, this small elasticity estimate for prime-aged males applies to labor supply at all ages, because as noted earlier and emphasized in Figure 2, a permanent change in work incentives leads to a proportional shift in the entire lifetime hours profile.

But now imagine that one interprets these results using the extensive margin model. When considering prime-aged males, it is natural to think that the relevant portion of Figure 3 is the middle segment in which employment is equal to one. As shown in Figure 3, the dashed and solid lines only differ at the edges of the life cycle, not in the middle part of the lifecycle. If one uses this model to guide empirical work, it would seem a particularly poor choice to focus on the response of prime-aged males. One would naturally expect a very small elasticity, but finding a small response for this particular group has effectively no information about the overall lifetime effect.

Put somewhat differently, viewed from the perspective of our extensive margin model, a small or even zero elasticity for prime-aged males does not tell us anything about the overall response of lifetime labor supply. Hence, it does not tell us anything about the elasticity of aggregate labor supply.<sup>9</sup>

### **A Hybrid Model**

The two models considered above are extreme: one features all adjustment along the intensive margin and the other features all adjustment along the extensive margin. Reality seems likely to lie somewhere in between, with some adjustment along both margins, and later on I will present data to show this. Here, I present a hybrid model that allows for adjustment along both margins.

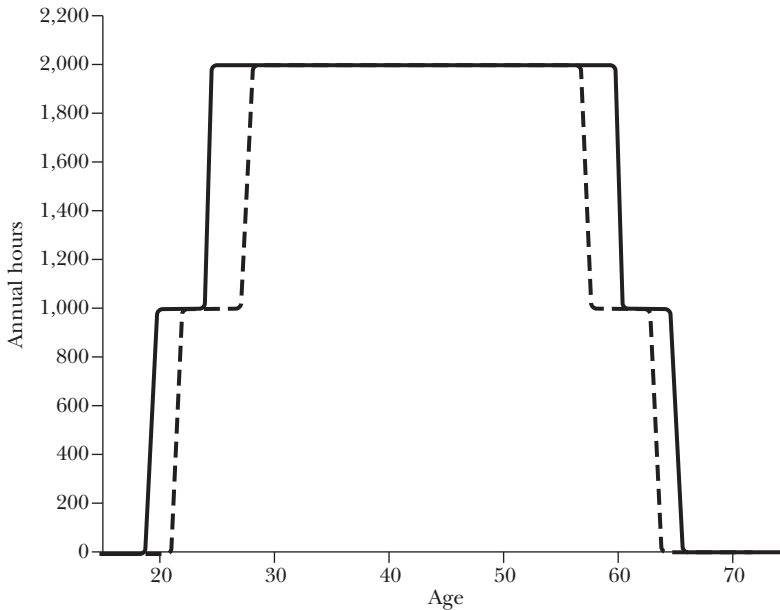
For this model, we expand the choice set for a given individual so that at each point in time they can choose between full-time work, part-time work, and no work. Conditional on full-time work, the hours are always the same, and conditional on part-time work, the hours are always the same. In this setting, an individual will optimally divide periods into three groups: one group for which they work full time, another group for which they work part time, and a third group for which they do not work. As in the extensive margin model, in any given period this will depend on both the wage rate of the worker and the disutility that they attach to work. Periods of full-time work will be those when wages are relatively high and/or disutility from work is relatively low, and periods of no work will be those with relatively low wages and/or high disutility of work. Generalizing our previous analysis, a permanent increase in the size of the tax and transfer system in this model will lead to some periods of full-time work potentially switching to become periods of part-time work,

hours of work for a relatively large group of individuals. MaCurdy (1981), and Altonji (1986) are two of the early and influential papers that used these data to study labor supply elasticities.

<sup>9</sup>My discussion focuses on how incorporation of the extensive margin affects the significance of small prime age male elasticities for the aggregate labor supply elasticities. But I do want to note that some recent work argues that elasticities for prime-aged males may also be quite a bit larger than found in earlier studies. See Keane (2011) and Keane and Rogerson (2012) for further discussion.



Figure 4

**Lifetime Hours in the Hybrid Model**

Source: Author's creation.

Note: Illustration of life cycle labor supply in hybrid model. Dashed line shows the effect of an increase in the size of the tax and transfer system.

and for some periods of part-time work potentially switching to not working. If an individual switches from full time work to part time work, this will show up as an adjustment along the intensive margin.

Figure 4 illustrates how this model compares with the extensive-margin-only model.

The key feature of this figure is that when the tax and transfer system is expanded, all of the boundaries shift in. The period of full-time work shrinks, and the period of no work expands. Some periods in which work was originally full time now become part time, and some periods that were originally part time become periods of not working. An important distinction from the intensive-margin-only model illustrated in Figure 2 is that in this model the adjustment along the intensive margin is also age-specific.

### **Relating the Extensive and Intensive Margins to the Data**

Earlier in this paper we constructed the data for hours worked per person by taking the product of the employment-to-population ratio and average hours per worker. Because the employment-to-population ratio measures the extensive

Table 3

**Cross Country Differences along Intensive and Extensive Margins**

<i>Panel A. Extensive margin: Employment-to-population ratio</i>				
<50	[50, 55]	[55, 60]	[60, 65]	>65
Greece (40.9)	Belgium (50.0)	Ireland (57.7)	US (60.1)	Switzerland (65.1)
Italy (44.1)	France (50.5)	Austria (57.9)	UK (60.4)	N. Zealand (66.9)
Spain (48.6)	Portugal (52.3)	Denmark (58.3)	Korea (60.7)	Sweden (67.6)
	Finland (54.4)	Germany (58.8)	Australia (61.7)	
		Japan (59.0)	Canada (61.8)	
			Netherlands (62.3)	
<i>Panel B. Intensive margin: Annual hours worked</i>				
<1,500	[1,500, 1,650]	[1,650, 1,750]	>1,750	
Germany (1,388)	Austria (1,502)	Spain (1,693)	New Zealand (1,761)	
Denmark (1,395)	France (1,516)	Japan (1,693)	US (1,825)	
Norway (1,423)	UK (1,535)	Canada (1,699)	Greece (1,941)	
Netherlands (1,435)	Finland (1,549)	Italy (1,718)	Korea (2,026)	
Sweden (1,466)	Switzerland (1,563)	Ireland (1,720)		
	Belgium (1,577)	Portugal (1,736)		
		Australia (1,737)		

Source: Panel A from OECD (2024a). Panel B from OECD (2024c).

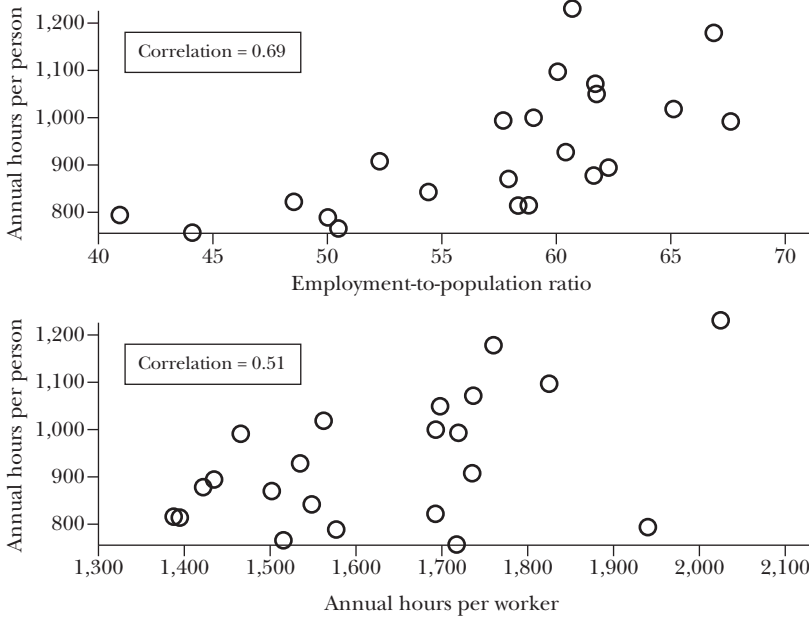
Note: Tables show averages for 2015–2019.

margin, and average hours per worker measures the intensive margin, reporting the two values separately allows us to gauge the importance of the two margins. The two panels of Table 3 present the results, once again representing averages for the period 2015–2019.

Studying the intensive and extensive margins separately reveals two interesting patterns. First, there is substantial variation across both margins. Variation along the extensive margin of employment-to-population ratios ranges from a low of 40.9 percent in Greece to a high of 67.6 percent in Sweden. (The United States has an employment-to-population ratio of 60.1 percent.) Variation along the intensive margin of average hours per worker varies from a low of 1,388 hours in Germany to a high of 2,026 in Korea. (The value in the United States is 1,825 hours per worker.) Second, the ranking of countries along the two margins can be very different. For example, whereas Greece, Italy, and Spain have the three lowest values for the employment-to-population ratio, they each have reasonably high values for average hours per worker. Conversely, while the Netherlands and Sweden have relatively high values for the employment-to-population ratio, they each have relatively low values for hours per worker. In fact, the correlation between the two measures across countries is modestly negative, at -0.27.

Although the two measures are modestly negatively correlated with each other, it is nonetheless true that each of them is quite strongly positively correlated

Figure 5  
**Total Hours, Hours per Worker, and Employment**



Source: Author’s calculation using data from OECD (2024a, c).  
 Note: Each point is a country. Values are averages for 2015–2019.

with total hours per person. Figure 5 illustrates these correlations across the 22 countries.

Having established that both margins are positively correlated with total hours per person, one might ask which of the two margins is more important in accounting for differences in hours per person. Since average hours per person is the product of the employment-to-population ratio and the average hours per worker, taking logs of this equation we have:

$$\log(\text{hours per person}) = \log(\text{employment ratio}) + \log(\text{hours per worker}).$$

A natural way to measure the amount of dispersion in a particular outcome is to compute its variance. Computing the variance of the log of the employment-to-population ratio and the variance of the log of average hours per worker provides a measure of how much each of the two measures contributes to the variance of log hours per person.

Doing this calculation, one finds that the variance of log hours per person is equal to 0.020, the variance of the log of the employment ratio is 0.017 and the variance of the log of average hours per worker is 0.011. (The sum of the two

variances is greater than the variance of the log of hours per person because the two measures are weakly negative correlated.) In other words, the variance in the log of the employment-to-population ratio is roughly 1.5 times larger than the variance of log hours per worker.

Consistent with the hybrid model that we introduced in the previous section, some of the dispersion in hours per worker is accounted for by variation in the fraction of full- versus part-time workers. The OECD also provides data on the fraction of employment that represents full-time work, and this variable has a correlation of 0.41 with average hours per worker. Another source of differences in annual hours per worker is variation in number of vacation days and statutory holidays.

Earlier we showed that there was a strong negative correlation relation between hours per person and government revenues as a percentage of GDP. This relationship continues to hold when we consider the intensive and extensive margins separately, as shown in the two panels of Figure 6.

A key message from the hybrid model (as well as from the extensive-margin-only model) is that one should expect employment-to-population variation to be very different across age groups. To examine this I consider the dispersion in the log of the employment-to-population rate across the 22 countries for three different age groups.<sup>10</sup> Table 3 reports these results.

The first entry reports the dispersion for the aggregate employment rate, while the next three columns show how dispersion varies across the life cycle. Motivated by our previous discussion, we consider three age groups: one in the middle of the life cycle and one at each of the “edges” where entry and exit from employment are most prevalent. A clear pattern results: dispersion is highest for the youngest age group and lowest for the middle age group. Dispersion for the middle age group is substantially smaller than overall dispersion, and dispersion for the old and young groups are dramatically higher than overall dispersion. The magnitude of the differences across age groups is dramatic: the variance for the oldest age group is an order of magnitude larger than for the middle age group, and the variance for the youngest age group is almost two orders of magnitude larger than the variance for the middle age group. This is exactly the pattern predicted by the extensive-margin-only model, in which employment rate responses to differences in work incentives are concentrated in young and older age groups.<sup>11</sup>

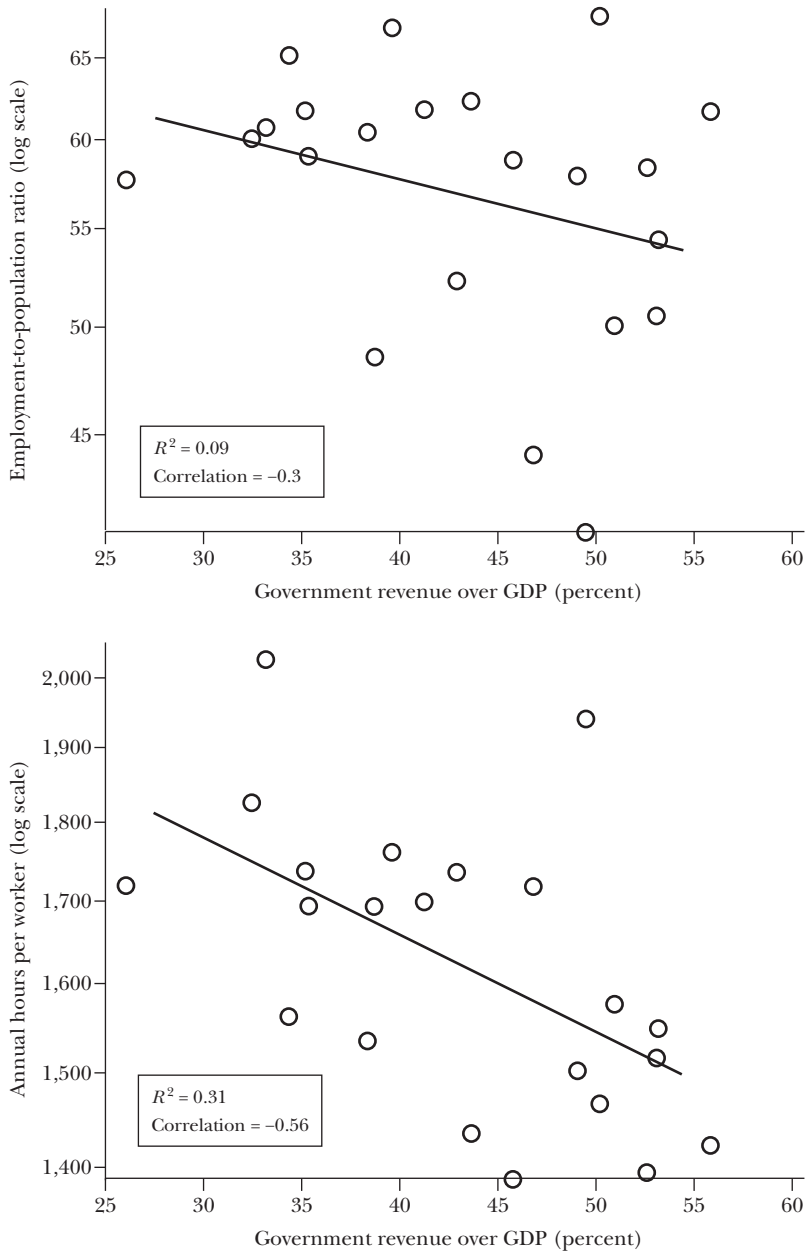
Another striking fact is that the employment rates for each of the groups shown in Table 3 are very highly correlated with the overall employment rate. That is, even though there is very little variation in the employment rates for prime age individuals across the 22 countries in our sample, the variation for this group is still highly correlated with the variation in the aggregate employment rate. The

<sup>10</sup>It is also of interest to study how differences along the intensive margin vary with age, but the OECD data that I am using do not report hours per worker by age group.

<sup>11</sup>The logic of the hybrid model also has important implications for differences in labor supply responses for married males and females, because on average females still have lower wages than males, and thus are more likely to be the marginal worker in two member households. Pursuing this issue is beyond the scope of the analysis here, but the reader can find some discussion of this in Keane and Rogerson (2015).

Figure 6

## Size of Government and Work along the Intensive and Extensive Margins



Source: Author's calculation using data from OECD (2024a, b) for top panel and OECD (2024b, c) for bottom panel.

Note: Each point is a country. Values are averages for 2015–2019.

Table 4

**Variance of the log Employment-to-Population Rate by Age Group**

15+	15–24	35–44	55–64
0.0169	0.1957	0.0029	0.0312

Source: OECD (2024a).

Note: Employment-to-population ratio is the average over the period 2015–2019.

correlation between the employment rate for prime aged individuals and the overall employment rate is 0.69, while the correlations for the 15–24 group and the 55–64 group with the overall employment rate are 0.85 and 0.86 respectively. The correlation between the employment rates for the young and old groups is 0.68.

While one might imagine that the factors shaping outcomes for young and older workers are radically different, even these outcomes are very highly correlated across countries. This pattern of high correlations is consistent with an explanation in which a common factor is behind the variation in employment rates for each of these groups, but that the dispersion in each group responds differently to this underlying common factor. As noted earlier, this message is in sharp contrast to the implications from the intensive-margin-only model that we described in the previous section.

## Conclusion

Hours of work differ substantially across advanced countries, and the magnitude of these differences implies large effects on output. In this paper, I have argued that a significant part of these cross-country differences may reflect labor supply responses to policies that differ across countries. An important element of this argument is that the aggregate labor supply elasticity is likely much larger than once thought. Previous research has implicitly assumed that small intensive-margin elasticities for prime-aged males necessarily translate into small overall aggregate elasticities. By examining the economics of labor supply along the intensive and extensive margin, I have argued that a small intensive margin for prime-aged males has little bearing on the magnitude of the aggregate labor supply elasticity.

The data show quite clearly that the extensive margin is an important source of variation in hours, and responses along the extensive margin vary dramatically across age groups. Viewed through the lens of a model that incorporates both an intensive and extensive margin, the patterns in the data are consistent with a large aggregate labor supply elasticity. One implication of this analysis is that labor supply matters for understanding macroeconomic outcomes across countries, and that undergraduate macro textbooks should include labor supply as an important issue for understanding macroeconomic outcomes.

Recent empirical studies that seek to estimate aggregate labor supply elasticities have incorporated choice along the extensive margin into their analyses, and are finding relatively large aggregate labor elasticities (for an overview of this research, see Keane and Rogerson 2015). While our understanding of the determinants and estimation of the aggregate labor supply elasticity remains incomplete, these estimates point in the direction of labor supply being an important element to be included in macroeconomic analyses.

## References

- Acemoglu, Daron, and Robert Shimer.** 1999. "Efficient Unemployment Insurance." *Journal of Political Economy* 107 (5): 893–927.
- Altonji, Joseph G.** 1986. "Intertemporal Substitution in Labor Supply: Evidence from Micro Data." *Journal of Political Economy* 94 (3): S176–215.
- Fernández, Raquel, and Alessandra Fogli.** 2009. "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics* 1 (1): 146–77.
- Keane, Michael P.** 2011. "Labor Supply and Taxes: A Survey." *Journal of Economic Literature* 49 (4): 961–1075.
- Keane, Michael, and Richard Rogerson.** 2012. "Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom." *Journal of Economic Literature* 50 (2): 464–76.
- Keane, Michael, and Richard Rogerson.** 2015. "Reconciling Micro and Macro Labor Supply Elasticities: A Structural Perspective." *Annual Review of Economics* 7: 89–117.
- Kleven, Henrik Jacobsen.** 2014. "How Can Scandinavia Tax So Much?" *Journal of Economic Perspectives* 28 (4): 77–98.
- MaCurdy, Thomas E.** 1981. "An Empirical Model of Labor Supply in a Life-Cycle Setting." *Journal of Political Economy* 89 (6): 1059–85.
- OECD.** 2024a. "Labour Market Statistics: Labour Force Statistics by Sex and Age." OECD Employment and Labour Market Statistics. <https://doi.org/10.1787/data-00309-en> (accessed April 21, 2024).
- OECD.** 2024b. "National Accounts at a Glance." OECD National Accounts Statistics. <https://doi.org/10.1787/data-00369-en> (accessed April 21, 2024).
- OECD.** 2024c. "Hours Worked: Average Annual Hours Actually Worked." OECD Employment and Labour Market Statistics. <https://doi.org/10.1787/data-00303-en> (accessed April 21, 2024).
- OECD.** 2024d. "Labour Market Statistics: Full-Time Part-Time Employment—Common Definition." OECD Employment and Labour Market Statistics. <https://doi.org/10.1787/data-00298-en> (accessed February 25, 2024).
- Ohanian, Lee, Andrea Raffo, and Richard Rogerson.** 2008. "Long-Term Changes in Labor Supply and Taxes: Evidence from OECD Countries, 1956–2004." *Journal of Monetary Economics* 55 (8): 1353–62.
- Rogerson, Richard.** 2007. "Taxation and Market Work: Is Scandinavia an Outlier?" *Economic Theory* 32: 59–85.
- Rogerson, Richard.** 2024. "Replication data for: Why Labor Supply Matters for Macroeconomics." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E198929V1>.
- Saez, Emmanuel, Joel Slemrod, and Seth H. Giertz.** 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature* 50 (1): 3–50.





## How Cyclical Is the User Cost of Labor?

Marianna Kudlyak

**M**acroeconomists have long been interested in the cyclical behavior of real wages, as part of the task of disentangling sources of employment fluctuations during the business cycle. For example, in the second half of the 1930s, John Maynard Keynes argued (in agreement with the earlier generation of neoclassical economists) that economic expansions tended to drive down real wages, because in an expansion the price level tended to rise more quickly than nominal wages. In his view, real wages were countercyclical. On the other side, John Dunlop and Lorie Tarshis carried out empirical analysis suggesting that real wages instead were procyclical—a finding that Keynes seemed open to accepting (for an overview of the debate in this journal, see Dunlop 1998).

To understand what is at stake, consider a textbook model of the aggregate labor market with a downward-sloping labor demand and an upward-sloping labor supply curve. At an initial equilibrium, labor demand equals labor supply at some wage and some level of employment. We know that the quantity of labor falls during a recession. In one class of models, a cyclical decline in employment happens from a shift to the left in labor supply along the stable labor demand curve and is accompanied

■ *Marianna Kudlyak is a Research Advisor (Economist and Bank Officer), Federal Reserve Bank of San Francisco, San Francisco, California. She is also a Visiting Fellow, Hoover Institution at Stanford University, Stanford, California; Research Fellow, Centre for Economic Policy Research (CEPR), London, United Kingdom; and Research Fellow, Institute of Labor Economics (IZA), Bonn, Germany. Her email address is [marianna.kudlyak@sf.frb.org](mailto:marianna.kudlyak@sf.frb.org).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.2.159>.

by an increase in wages. The result is a Keynesian countercyclical real wage, in which a reduction in employment in recessions is accompanied by a rise in real wages. In a different class of models, the cyclical decline in quantity of employment happens via a shift of the labor demand curve to the left. In this case, a recession is accompanied by a decline in wages—that is, a procyclical real wage. Overall, the quantity of labor could decline as a result of a shift in either labor demand or labor supply—but these explanations have different implications for the cyclicalities of real wages. Economists measure the cyclicalities of a given data series by the degree to which it co-moves with a measure of the business cycle, for example, a (detrended) unemployment rate.

In many standard macroeconomics models, a rigid (or countercyclical) wage is the key to generating employment fluctuations. If, instead, real wage is strongly procyclical—as I will argue in this paper—one needs to look for other factors to explain the large declines in employment that happen in recessions.

In the next section, I discuss an earlier literature, which took the average real wage as the measure of real marginal cost of labor and found it almost acyclical. However, an aggregate wage can change for two reasons during the business cycle: an actual change in the wage and/or a shift in the composition of the labor force. If one adjusts for the fact that lower-wage workers are more likely to exit the labor force in a recession, the resulting composition-adjusted wage is procyclical. A further distinction between wages of incumbent workers and wages of newly hired workers, shows that wages of newly hired workers are considerably more procyclical than the average wage. Each of these subsequent refinements of wage is more procyclical than an average wage.

The rest of the paper explains why the observed wage does not capture the appropriate economic concept, which is the real marginal cost of labor, and how the price of labor can be much more procyclical than the average wage or even the new hire wage. In long-term labor relationships, neither an average wage nor the new hire wage fully captures the price of labor. As Hall (1980, p. 101) writes: “[T]o see what is happening today in the labor market, one should look at the implicit asset prices of labor contracts recently negotiated, not at the average rate of compensation paid to all workers.” The expected long-term nature of the relationship allows the firm to smooth some potential fluctuations in worker wages over time. Therefore, the current wage does not capture the contemporaneous cost of a worker to a firm. I will introduce the concept of the user cost of labor (Kudlyak 2014), an analog to the user cost of capital, as a way to measure a price of any long-term-use asset. In this view, the wage is an installment payment on an implicit long-term agreement between a worker and a firm. I argue that the user cost of labor, not a per period wage, plays a role in firm hiring decisions. Indeed, it is quite possible to have average wages that appear rigid during the business cycle and also to have a user cost of labor that is highly procyclical. I also discuss a measurement issue associated with cyclical changes in labor’s price: the possibility of confounding effects of cyclical variation in the quality of labor matches created at different points during the business cycle, an issue distinct from the composition effect mentioned above.

With these conceptual building blocks in place for how to think about the price of labor, I describe recent estimates of the cyclical price of labor as estimated by the cyclical price of the user cost. Recent work estimates that the price of labor is strongly procyclical, declining by more than 4.5 percent when unemployment increases by 1 percentage point (Bils, Kudlyak, and Lins 2023). This evidence of a strongly procyclical price of labor raises open research questions of how to model labor markets in a way that leads to the observed variations in employment over the business cycle.

### **Cyclical Price of the Average Wage, Wages of Workers in Ongoing Relationships, and Wages of New Hires**

An early literature took the actual average real wage in the economy as the measure of real marginal cost of labor (for example, Mankiw, Rotemberg, and Summers 1985). The average wage is almost acyclical. Figure 1 shows big unemployment increases during recession but no corresponding declines in the average wage. This observation led to a conclusion that the price of labor over the business cycle was rigid; that is, it falls little or not at all during recessions and rises little or not at all during cyclical booms.

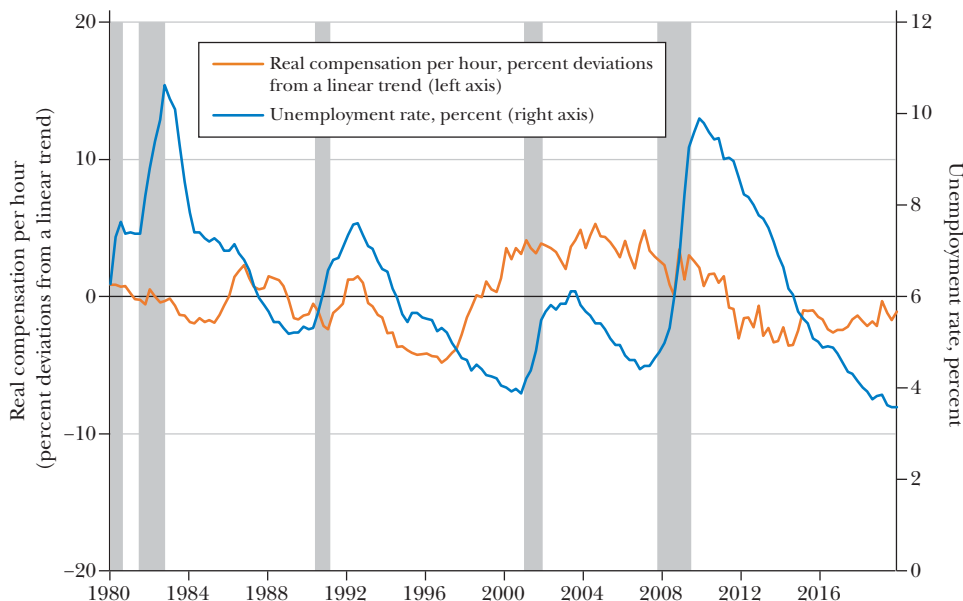
Changes of the average wage, however, reflect both the changes of wages of individual workers and also changes in the composition of the workforce. For example, recessions are times when low-wage workers tend to leave the workforce, while economic booms are the times when low-wage workers tend to be more widely represented in the workforce. Such countercyclical movements in the composition of workforce, in terms of their typical relative wages, will affect the average wage.

Research has sought to disentangle the cyclical price of wages from cyclical changes of the composition of the workforce.<sup>1</sup> Starting from Bils (1985), this literature turned to examining cyclical price of wages of individual workers, without aggregating. An analysis of the cyclical price of individual wages is conducted with panel data, which contain observations about a cross-section of workers over time. The individual-level panel data allow not only controlling for observable worker characteristics, such as education or age, but also for unobservable worker fixed effects. Using panel data on young men from 1966 to 1980, Bils finds real wages to be procyclical—a percentage point rise in the unemployment rate is associated with a decrease in real wages of between 1.5 and 2 percent.

More recently, the composition effect in wages is demonstrated in Cajner et al. (2020). Using weekly administrative payroll data from a large US payroll processing company, they show that during the 2020 pandemic recession, average (nominal) wages increased by nearly 6 percent through mid-May; however, the entire growth was attributed to the changing composition of workers (see their Figure 5).

<sup>1</sup>The argument about the composition bias is first mentioned in Stockman (1983). See also Solon, Barsky, and Parker (1994).

Figure 1

**Real Compensation per Hour and Unemployment Rate**

Source: Created using data from the US Bureau of Labor Statistics.

Note: The figure shows compensation per hour in the nonfarm business sector adjusted for inflation by changes in the CPI-U and the unemployment rate. The compensation series are expressed as In deviations from a linear trend. Quarterly, seasonally adjusted series, 1980–2019. For details, see the online Appendix.

Specifically, they show that at the onset of the pandemic, workers at the bottom of the wage distribution were much more likely to experience employment reductions than those at the top of the wage distribution. Consequently, from March through the end of April 2020, the workforce became more selected towards higher-wage workers, while the reverse happened thereafter.

In addition to estimating the cyclicity of individual wages, another marked advance of *Bils (1985)* is a distinction between wages of the job stayers who remain with the same employer versus wages of the job changers who are moving between employers or in and out of the workforce. *Bils* documents a significantly greater procyclicality of wages of job changers, relative to wages of job stayers. Specifically, as the unemployment rate goes up by one percentage point, wages of newly hired workers decrease by 3 percent, while wages in ongoing relationships decrease by less than 1 percent. This finding is of marked interest, as *Bils* notes, if these more transient workers are the ones who actually reflect cyclical changes in supply and demand in the labor market.

Bringing this evidence to the analysis of the canonical search and matching model of the labor market, *Pissarides (2009)* discusses how job creation is influenced

by wages in new matches. To the extent that the new hire wage captures the marginal price of labor, the relevant price of labor for job creation is more procyclical than what is implied by the cyclical nature of the average wage. However, as I argue below, even the new hire wage does not fully capture the price of labor.

## **Pricing Labor by Its User Cost**

### **History Dependence of Wages**

Employment relationships are often long-term. In a long-term relationship, the wage observed in a relationship—either the new-hire or the average wage—may not capture the true price of labor. Instead, in a long-term agreement, the current wage is a per-period payment within the context of an implicit agreement between a worker and a firm, and that payment does not necessarily equal the period's marginal cost of worker.

This insight is illustrated by a fact pattern described in the previous section: wages of newly hired workers are substantially more procyclical than wages of workers in ongoing relationships. Moreover, individual wages exhibit dependence on the history of economic conditions since the start of the match (for example, Beaudry and DiNardo 1991; von Wachter 2020). Taken together, these observations imply that the current wage alone does not summarize the wage commitment a firm makes upon hiring a worker. Adding a worker in one period, rather than in the following period, may affect not only the wage at the time of hiring, but also the future path of wages in the relationship.

The concept of the user cost of labor builds on the ideas of Barro (1977) and Hall (1980), among others, who argue that what matters to a firm is the value of wages to be paid during a firm-worker relationship. For example, Barro calls sticky wages just a “facade” of the implication of the long-term labor contracts during periods of short-term macro fluctuations. In the terms of Kydland and Prescott (1980), the weak procyclicality of real wage can suffer from “cyclical measurement bias” because, with implicit contracts, wage payments are not perfectly associated with labor services provided each period. Indeed, what is relevant for the volatility of job creation and unemployment is the rigidity of the present discounted value of wages that firms expect to pay to a worker at the time of hiring, extended over the course of the employment relationship (Shimer 2004).

### **The User Cost of Labor**

In employment relationships, labor is akin to long-term asset. The way to price a long-term asset, as we know from the study of physical capital, is by its rental price or user cost. The user cost of capital is the difference between the price of a unit of capital at the beginning of the period and the price at which the remainder of the unit, after depreciation, can be sold at the end of the period.

The user cost of labor, by analogy with the user cost of capital, is the difference between the expected present-discounted value of wages to be paid to a worker

hired at the beginning of a period and the expected present-discounted value of wages to be paid if this worker were hired the following period (Kudlyak 2014). In the case of labor, the appropriate discount rate includes an interest rate, but also the probability of job separation, which plays a role similar to capital depreciation.<sup>2</sup> Continuing the analogy of the user cost of capital, the user cost of labor reflects the value of “selling” the initial contract at the end of the period. The sale price of such a contract is the value of a new contract signed at the beginning of the following period.

The idea of the user cost of labor arises quite naturally when a firm evaluates the costs and benefits from adding labor.<sup>3</sup> Suppose that a firm-worker match can last indefinitely, but in each period it faces some exogenous probability of separation. Consider a firm deciding between hiring a worker now versus postponing the hiring until the following period. Under either decision, the firm will have an equal number of workers from the following period onward, however, there is an additional worker in this period if the decision is to hire now versus later. As a result, the benefit of hiring now versus later is having a worker in this period. Consequently, the cost of this decision represents a price of having an extra worker who works during this period—the price of labor. Because, as discussed above, hiring now versus later might have an impact on the entire wage path in the relationship, the cost of a decision to hire now is not simply the period hiring wage. Firms take into account the impact of hiring now on the path of future wages. The user cost of labor provides a way to capture such an impact.

Formally, to account for all the future costs and benefits of hiring now versus later, a firm compares (1) benefits and costs from a firm-worker match created now, versus (2) benefits and costs from a match created a period later. The costs include the expected present-discounted value of wages paid through the entire duration of the match as well as any match-creation costs, where the discounting takes into account a time discount factor as well as the probability of separation as mentioned above.

The benefit side of the firm’s decision to hire in period  $t$  versus later simplifies to equal the output in period  $t$ . This is because from  $t + 1$  onward, the output in a match that starts in  $t$  or in  $t + 1$  is the same. This is by design of the firm’s decision problem, whereby a firm considers hiring the same worker in  $t$  or in  $t + 1$ , whose expected output from  $t + 1$  onward is the same whether the match starts in either  $t$  or  $(t + 1)$ .

The cost side of this period- $t$  output is the user cost of labor. It consists of the wage component of the user cost of labor and the nonwage component of the user cost of labor. The nonwage component encompasses costs of searching and finding a good match, as well as training or any other costs associated with creating a match. (These costs are not necessarily only upfront and can have a more complicated term structure.)

<sup>2</sup>Labor’s depreciation can also be extended to include changes in labor’s productivity within a match.

<sup>3</sup>For formal derivations of this framework, see online Appendix A.

The wage component of the user cost of labor is the expected present discounted value of wages to be paid to a worker hired in period  $t$  minus the expected present discounted value of wages to be paid to a worker hired in  $t + 1$ . As discussed above, in general, wages depend on the history of the labor market conditions from the start of the match. For example, in  $t + 1$ , a worker hired in  $t$  might have a different wage than a worker hired in  $t + 1$ . Therefore, the future wages in the relationships that start in  $t$  versus in  $t + 1$  might differ.

If the separation rate and discount rate do not depend on the history from the start of the match, the expression for the wage component of the user cost of labor in period  $t$  simplifies to the sum of hiring wage in  $t$  plus the expected present discounted value of the differences from  $t + 1$  onward in wages in the match that starts in  $t$  and wages in the match that starts in  $t + 1$ . This second term captures the impact of hiring in  $t$  on future wages in the relationship. If wages do not depend on history of economic conditions from the start of the match, then the expected future wages in the two relationships are equal and the user cost equals the hiring wage—which in turn equals the period's average wage. This, however, is generally not the case. The expression of the user cost of labor can be extended to allow for history-dependent separation rates.<sup>4</sup>

## Which Wage Matters for Firm Hiring Decisions

### The Allocational Role of the Wage Component of the User Cost of Labor

Having introduced the user cost and, specifically, the wage component of the user cost of labor, I will demonstrate that it is the rigidity (or flexibility) of this statistic, and not of an average wage or of a new hire wage, that is relevant for firm decision-making about hiring.

From the discussion above, it follows that the firm's hiring decision involves comparing the output to the user cost of labor, and not to a one-period wage. That is, the wage component of the user cost of labor determines allocation.

An intuitive way to see the allocational role of the wage component of the user cost of labor is to extend the framework described above by adding a condition that the firm will hire a quantity of labor in a given period up until the value of the decision of hiring now versus later is driven to zero. Such a zero-profit condition is a standard assumption in many models. It ties the period's output with the period's

<sup>4</sup>With history dependence in separation rates, the expected number of workers in a match in any future period depends on when each of the matches started. Kudlyak (2014) and Bills, Kudlyak, and Lins (2023) derive an expression for the user cost of labor with history-dependent discount factor. With history-dependent separation rates, in addition to the hiring wage in  $t$  plus the expected present-discounted value of the differences from  $t + 1$  onward in wages in the match that starts in  $t$  and wages in the match that starts in  $t + 1$ , the wage component of the user costs of labor includes the expected present discounted value of wages to be paid in matches that start from period  $t + 2$  onward, weighted by the difference in the probabilities that there will be a need to create such a match to replace separated match in a position that starts in  $t$  versus in  $t + 1$ .

user cost of labor, which is the sum of the nonwage component and the wage component of the user cost of labor. The wage component is the allocational price of labor. At times, in this paper or elsewhere, we loosely refer to the wage component of the user cost of labor as simply the user cost, but it is the wage component that is the price of labor (in terms of wages).

Consider a textbook search and matching model (Mortensen and Pissarides 1994). It can be shown that in such models, the zero-profit condition described above corresponds to a free entry condition (Kudlyak 2014): in this case, the firm will post vacancies until the value of such a decision is driven to zero. The free entry condition equates the period's output to the sum of the wage component of the user cost and the nonwage component of the user cost, which in the model consists of the constant vacancy posting costs and the time-varying probability of filling a vacancy, which is a function of the vacancy-unemployment ratio. If the wage component of the user cost is rigid, then all changes in output must be channeled through changes in the nonwage component—specifically, through the probability of filling the vacancy via changes in vacancy-unemployment ratio. Furthermore, if the wage component of the user cost is rigid and constitutes a large fraction of output (for example, profit is small), then even small changes in output can translate into large changes in the vacancy-unemployment ratio. If, instead, the wage component of the user cost responds to shocks to the extent similar to the response of output, then one needs to think of other mechanisms to generate large changes in the vacancy-unemployment ratio observed in the data.

The question how responsive to shocks—in the present context, how cyclical—is the price of labor is an empirical one. Of course, economists have built models that deliver a rigid wage component of the user cost of labor using various mechanisms based on micro foundations. The key, however, is whether the rigidity of the wage component of the user cost of labor is supported in the data.

The concept of the wage component of the user cost of labor expresses the impact of the economic conditions at the time of hiring on future wages in terms of a current-period “flow value.”<sup>5</sup> From the conceptual point of view, measuring labor's price by its user cost provides a convenient concept of a flow price of labor, which can be readily contrasted with individual wages or other prices such as the flow value of nonwork, for example.<sup>6</sup> From the measurement perspective, as the sections below show, there are natural assumptions that allow setting the terms in the sum far into the future to zero, which simplifies the estimation.

<sup>5</sup>If the present-discounted value of wages to be paid in a match does not move over the business depending on when the match starts, then the wage component of the user cost of labor is also rigid. This is because the wage component of the user cost is simply the difference between the present-discounted values of wages paid in the relationships that start in the two consecutive periods. However, there are a few advantages of focusing on the cyclicity of the wage component of the user cost as opposed to on the cyclicity of the present discounted value of wages.

<sup>6</sup>Chodorow-Reich and Karabarbounis (2016) estimate that the flow opportunity cost of employment is procyclical and volatile over the business cycle.



### **How the Procyclicality of the Wage Component of the User Cost Can Be Masked by Seemingly Rigid Wages**

As described above, the wage component of the user cost is the sum of hiring wage in  $t$  plus the expected present-discounted value of the differences from  $t + 1$  onward in wages in the match that starts in  $t$  and wages in the match that starts in  $t + 1$ . Does accounting for these future terms—beyond the hiring wage—matter quantitatively for the estimation of the cyclical nature of the wage component of the user cost? In this section, I provide an intuition of how seemingly rigid wages within employment relationships can mask a strongly procyclical wage component of the user cost of labor. In the sections below, I will show that it indeed matters empirically.

Consider an example of a wage setting with history dependence: an insurance contract between a risk-averse worker and a risk-neutral firm (as in Thomas and Worrall 1988). Risk-averse workers dislike fluctuations in their wages. Thus, the firm insures the worker from business cycle fluctuations by offering a wage that remains fixed for as long as possible. For example, the firm and a worker might agree to split the joint surplus at the beginning of the match according to their respective bargaining weights and keep the wage constant until the value from the match to one of the parties becomes lower than that party's outside option. When the outside option binds, the wage in the relationship is adjusted just enough to prevent that party from leaving the match, provided the joint surplus is still positive.<sup>7</sup> Under such a wage-setting process, the wage under an ongoing contract is not necessarily equal to the wage that can be obtained under a newly signed contract.

In this setting, wages are smoothed—that is, wages are to some degree isolated from the business cycle fluctuations. In addition, wages depend on the history of labor market conditions at the start of the match. For example, the contracts that start during business cycle troughs, when the joint surplus is low, feature a lower stream of wages as compared to the contracts that start in economic booms, when the surplus is high. There is a lock-in to the initial business cycle conditions, at least for some period. Under such a wage setting with history dependence, wages of newly hired workers are procyclical, while wages of workers in ongoing relationships are shielded from the business cycle fluctuations and are less procyclical or even rigid.

Consider a firm hiring in a business cycle trough when the new hire wages are low. Once a worker is hired, the path of wages for that worker remains relatively low, at least for some time, as compared to a worker hired later during recovery. The wage component of the user cost of labor is even lower than the actual new hire wage, because it takes into account not only the low hiring wage but also the impact of the labor market conditions at the time of hiring on future wages. The opposite holds when a worker is hired in an economic boom—the wage component of the user cost of labor is higher than the new hire wage because it takes into account the

<sup>7</sup>Typically, this kind of model also includes a search and matching friction or an exclusion restriction from the labor market prevent either side from walking away from the contract and getting an equivalent of a spot market wage.

future higher wages in the relationship that starts in a boom. Consequently, if wages of workers in ongoing relationships are rigid and wages of newly hired workers are procyclical, we can expect the wage component of the user cost of labor to be even more procyclical than wages of newly hired workers.

Although this example is motivated by a model of implicit insurance contracts, the relevance of the user cost as a measure of price of labor does not hinge on the source of history dependence in wages. As another example, in macroeconomic models it is common to use a Calvo (1983) approach in which prices can only be changed at certain intervals, and so price adjustments in response to a shock do not happen all at once but are spread out over time. Calvo price-setting is history dependent and can be applied not just to labor markets but to prices across an economy, but it is not motivated by worker-firm insurance. Furthermore, the price of labor as measured by the user cost nests the case of wages with no history dependence.

Summarizing, if worker's wages exhibit substantial history dependence on the labor market conditions from the start of the match but are otherwise rigid, then a price of labor can be much more volatile than the new hire wage—much lower than the new hire wage if a worker is hired in a bust, and much higher than the new hire wage if a worker is hired in a boom. That is, rigid wage within employment relationships combined with a procyclical wage at the time of hiring can result in a procyclical present-discounted value of wages to be paid in the match, and, therefore, a procyclical wage component of the user cost of labor.

### **Different Methods of Wage Setting but the Same Volatility of the Wage Component of the User Cost—A Demonstration of Allocational Role of the Wage Component of the User Cost**

To illustrate how the wage component of the user cost of labor, and not wage, determines firm decisions about hiring, in Kudlyak (2014) I study a textbook search and matching model with four different methods of wage setting. My goal was to study wage settings that, by design, allow for different degrees of observed wage rigidity, from completely rigid wages within a match to wages renegotiated each period.

I embed the self-enforcing worker-firm insurance contract of Thomas and Worrall (1988) in a search and matching model with risk-averse workers and risk-neutral firms. I distinguish three types of contracts based on the degree of commitment to the contract: full commitment contracts from both the firm and the worker; contracts with lack of commitment from the worker's side and full commitment from the firm's side; and contracts with lack of commitment from both the worker's and firm's sides. In the optimal contract with commitment, the wage remains constant within a match. In the contracts with lack of commitment, the wage remains constant until the value of the outside option for the party without commitment exceeds the value under the contract, in which case the wage is adjusted to prevent renegotiating. For comparison, I also study fourth method of wage setting that involves Nash bargaining each period, in both new and existing job matches, as in the textbook search and matching model. I solve each of the

four models and obtain the series of the wage component of the user cost of labor, wages of newly hired workers, wages of workers in ongoing relationships, and the vacancy-unemployment ratio. I then calibrate models' parameters such that each model delivers the same targeted cyclicity of the wage component of the user cost.

The simulation results from the four models show that once the cyclicity of the wage component of the user cost is calibrated to the same number across the models with different wage settings, the models generate similar volatility of the vacancy-unemployment ratio.<sup>8</sup> That is, the volatility of the wage component of the user cost of labor controls allocations in the models.

However, even though the cyclicity of the wage component of the user cost is calibrated to the same number across the models, cyclicity of individual wages and the wages of newly hired workers differ significantly across the alternative models of wage setting. Specifically, in the three implicit insurance contract models, wages of newly hired workers are more procyclical than wages of workers in existing relationships. Average wages are least cyclical in the model with full commitment, more procyclical in the model with one-sided lack of commitment, and even more procyclical in the model with two-sided lack of commitment. In the model with full commitment, wages within employment relationships are rigid by assumption; and, therefore, the cyclicity of the average wage is due to the cyclicity of wages of new hires entering employment relationships. In the model with commitment on the firm's side and lack of commitment on the worker's side, in addition to the new hire wage cyclicity, the wages in the existing employment relationships are bid up whenever the worker's outside option value becomes more attractive than the value from the contract. In the model with lack of commitment on both the firm's and worker's sides, the wages can also be bid down whenever the value from the match for a firm falls below zero.

In the model with Nash bargaining each period in both new and existing matches, the wage component of the user cost of labor equals the new hire wage, which in turn equals the average wage. Consequently, under this wage setting, all three wage statistics—the new hire wage, the average wage, and the user cost of labor—share the same cyclicity.

Clearly, the behavior of individual wages in the models is a facade. The allocations are determined by the behavior of the wage component of the user cost of labor.

### **Isolating the Cyclical Variation in the Labor's Price from the Cyclical Variation in Match Quality**

One conceptual challenge in estimating the cyclicity of the price of labor from the data is distinguishing the true cyclical change in the price from the cyclical

<sup>8</sup>There are slight differences due to the curvature of the utility function with the risk-averse preferences of workers.

differences in the productivity of matches created at different points in the business cycle. Specifically, matches created at different points in the business cycle might systematically differ by worker productivity, firm productivity, or the interaction of firm–worker productivity. We refer to this productivity as “match quality.” This variation in match quality differs from the cyclical variation in the composition of workforce described earlier. Variation in the composition of the workforce typically refers to the characteristics of a worker that can be controlled for by observables, like age or education. In contrast, match quality of the same worker might differ if the worker is matched with one firm versus with another, and can be a function of both the worker and the firm characteristics.

In estimating the cyclical component of the user cost of labor, we want to isolate the changes in the price of labor from any changes in match quality. The wage component of the user cost of labor is not directly observed in the data, because researchers observe only an actual path of wages of the hired worker and not a hypothetical expected wage path of the same worker if that person was hired in the following period. These expected would-be wage paths need to be inferred from the observed wage paths of the workers hired the following period. However, the matches created the following period might be of different quality than the matches created in the current period.

Note that the problem of disentangling the cyclical component of the price from the cyclical component of match quality is not specific to the estimation of the wage component of the user cost of labor; it also arises in estimating the cyclical component of wages of newly hired workers. Specifically, to estimate the latter we would like to measure what the change in the new hire age will be if the same worker was hired at different points over the business cycle. However, in practice, we have data on wages of newly hired workers from different matches created in different periods over the business cycle and, potentially, of different quality.

Not accounting for the cyclical variation in quality of the matches can lead to a bias in the estimates of the cyclical component. For example, if the quality of matches created in recessions is typically higher than the quality of matches created in booms (that is, if match quality is countercyclical), the higher quality will be reflected in relatively higher wage paths of the matches created in recessions. If not accounted for, this would bias the estimates of the cyclical component of the price of labor countercyclically, so that estimates would show wages to be less procyclical than they really are.

A priori, it is not obvious whether match quality is pro- or countercyclical. For example, a greater availability of potential hires during recessions might increase the quality threshold for new matches, leading to creation of better matches and a countercyclical match quality. This view is consistent with what has been called “cleansing effect” of recessions (Caballero and Hammour 1994). Alternatively, matches created in recessions might be of a worse quality, perhaps because employers are more likely to be drawing on unemployed workers whose skills have depreciated. In that case recessions are sully, and match quality is procyclical (for a model of sullying recessions, see Barlevy 2002). Ultimately, the cyclical component of match quality is an empirical question. Here, I describe two widely used approaches

and a new approach of Bilts, Kudlyak, and Lins (2023) of how to control for match quality.

### **Approach #1: Wage Changes**

One of the widely used approaches of controlling for match quality in estimating the cyclical of wages is based on looking at the cyclical of wage *changes* (for example, Bilts 1985; Gertler, Huckfeldt, and Trigari 2020). Using individual worker panel data, the researcher calculates the change of a worker's wage between the past period and the current period and estimates the cyclical of this change. If this change is taken for wages within an employment relationship, the differencing takes out any match-specific fixed effects. (Such match-specific effects—fixed for the short- or the medium-run—within a match represent a typical concern in the match quality literature.)

When the wage change approach is applied to the new hire wage, in that case, typically, the change is calculated as the difference between the new hire wage in the current match and the most recent wage of the same worker, which is the last wage in the previous match that the worker was in (possibly prior to an intervening nonemployment spell). Two concerns arise here: first, the previous wage is not a new hire wage and is subject to all kinds of wage smoothing issues discussed above; and second, most importantly, this approach implicitly uses the quality from the previous match as a proxy for the quality of the current match. Bilts, Kudlyak, and Lins (2023) provide a discussion of these issues in details.

In this approach, the estimate of the cyclical of the new hire wage will be biased if the change in the quality between the two consecutive matches is correlated with the business cycle. If workers tend to move to better-quality matches during economic recoveries (when unemployment falls), then the estimate of the cyclical of new hire wages will be biased procyclically. This might be the case for the job-to-job switchers, who typically move to better matches during recoveries. If, instead, workers move to lower-quality matches during recoveries, then the estimate of the cyclical of new hire wages will be biased countercyclically. This might be the case for the new hires from unemployment during recoveries. These workers, having separated from their jobs during recessions, typically fall off the job ladder, search for jobs, and start climbing the job ladder anew.

### **Approach #2: Fixed Effects**

Another widely used approach of controlling for match quality in estimating the cyclical of wages is to control for individual worker fixed effects (for example, Kudlyak 2014). Panel data on individual workers typically allow identifying worker fixed effects from observations on the same worker employed at different firms over time.

A worker fixed effect captures the average quality of all matches in which the worker in the panel has been observed. Implicitly, that average quality serves as a proxy for the quality of the current match. The intuition for the direction of the bias from this approach is similar to the one from the wage change approach. To

the extent that during recoveries workers move to better than their average-quality matches, this will bias the estimates of the new-hire wage cyclicality procyclically. And, conversely, to the extent that during recoveries workers move to worse than their average quality matches, this will bias the estimates of the new-hire wage cyclicality countercyclically.

If a researcher has access not only to panel data but also to matched employer-employee data, such data allow controlling for both firm and worker fixed effects. These fixed effects are identified from observing the same worker employed at different firms and a firm employing different workers over time, respectively. These separate worker and firm fixed effects might not fully capture a joint effect of worker–firm match quality. However, using a joint worker–firm match fixed effect for estimating the cyclicality of the new hire wage leads to the same biases as using simply a worker fixed effects—the fixed effects are calculated as an average over different matches.

### **Approach #3: The Match’s Long-Run Wage**

The long-run wage in the same match can be used as a proxy for the new hire match quality (Bellou and Kaymak 2021; Bills, Kudlyak, and Lins 2023). Under this approach, the long-run wage in the same match is differenced out from the new hire wage. Such differencing takes care of any quality of a match that is fixed throughout the match—which is a concern of most of the match-quality literature.

The issue may arise if the quality of a match evolves during the match and its evolution is correlated with the economic conditions at the start of the match. Specifically, if quality of the match grows during the course of the match *and*, importantly, this growth is greater in the matches that start in recessions versus booms, the estimates of the cyclicality of the new hire wage might be biased procyclically. Bills, Kudlyak, and Lins (2023) perform multiple robustness exercises to check whether matches that start in recessions display disproportionately greater growth of quality within the match, either fundamentally or via selection in the matches that survive to their definition of the long run of eight years, as compared to the matches that start in booms. They find only mild evidence of selection.

## **Estimates of the Cyclicity of the Price of Labor**

Up to this point, we have focused on the conceptual issues involved in estimating and interpreting the user cost of labor: thinking about implicit long-term contracts, where wages are smoothed and the economic conditions at the time of hiring (say, during a recession or a recovery) may have an impact on the entire future path of wages in the relationship, and also thinking about a need to adjust for the possibility that matches created at different points of the business cycle can be of different quality. What happens when you put all these ingredients together and carry out an actual estimate?

Bils, Kudlyak, and Lins (2023) estimate the cyclicity of new hire wages and of the wage component of the user cost using individual worker wage data from the National Longitudinal Surveys, the NLSY1979 and NLSY1997, spanning 1980 to 2019. They construct the empirical counterpart of the wage component of the user cost based on the behavior of individual wages and turnover.

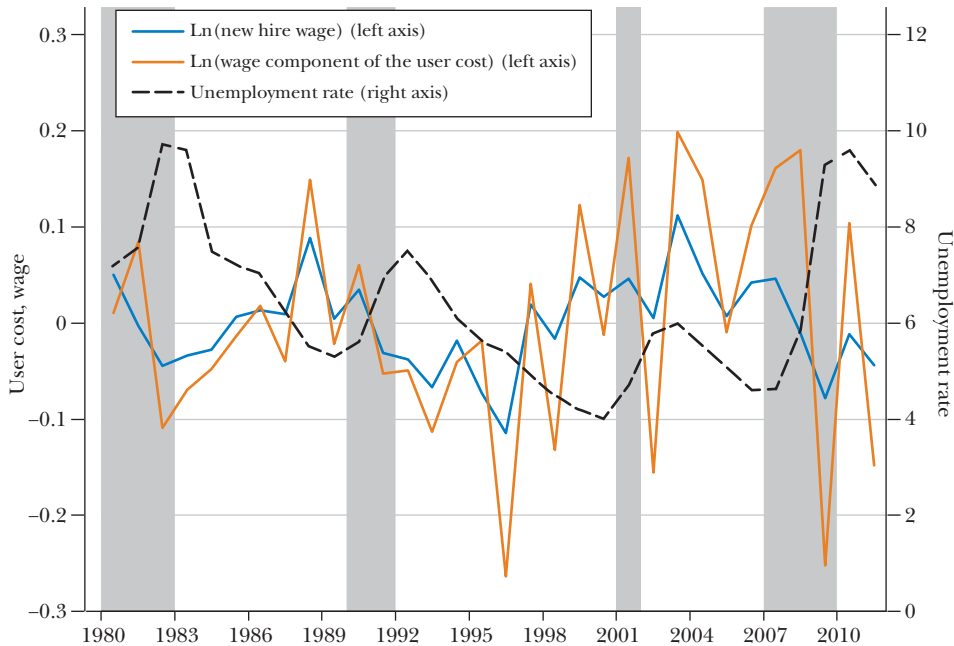
First, they construct a measure of the real wages for new hires after doing an adjustment for match quality, using (as discussed above) the expected long-run wage of a match as proxy for firm–worker match quality. Second, for each year in their sample period, they calculate the job separation rate conditional on the year when the match starts. Third, using the quality-adjusted wages and the job separation rates, they construct 32 annual observations of the wage component of the user cost of labor, from 1980 to 2011. Each observation involves calculating the expected wage path eight years out, and they explore a variety of assumptions about time variation in the separation rate and the appropriate time discount factor. Fourth, they investigate whether this measure of the user cost is cyclical, by regressing the log of the user cost of labor on the unemployment rate and a cubic trend. Figure 2 shows the quality-adjusted new-hire wage and the wage component of the user cost of labor with time-varying and history-dependent separation rate and discount factor from Bils, Kudlyak, and Lins (2023).

In the Bils, Kudlyak, and Lins (2023) estimates, the quality-adjusted wage of new hires decreases by 2.35 percent when unemployment goes up by 1 percentage point. It is nearly as cyclical for hires from nonemployment as for those moving job-to-job. (This finding, like all others mentioned here, is statistically significant.)

The next step for Bils, Kudlyak, and Lins (2023) is to calculate the wage component of the user cost of labor, under various assumptions of the time variation in the separation rates and discount factors, and estimate its cyclicity. Under all these various assumptions they find that the wage component of the user cost of labor is strongly procyclical. Specifically, with a constant separation rate and discount rate, a percentage point higher unemployment rate is associated with a decline of  $-4.81$  percent in the wage component of user cost. With the separation rates that vary both with the current year and the match’s start year and the time-varying discount factor, their benchmark estimate, the wage component of the user cost of labor declines by  $-5.32$  percent when the unemployment rate increases by 1 percentage point. Its estimated elasticity with respect to real GDP is around 2.6.

Finally, Bils, Kudlyak, and Lins (2023) note that that matches created in recessions typically last shorter than the matches created in booms—that is, the durability of matches by the match start date is procyclical (see, among others, Bowlus 1995; Mustre-Del-Rio 2019; Baydur and Mukoyama 2020). If the new hires that start in recessions display systematically higher separation rates, then firms that hire in recessions will eventually need to spend more on hiring to create replacement matches. Such higher expected turnover costs might be reflected in the lower wages of matches that start in recessions. Therefore, Bils, Kudlyak, and Lins (2023) calculate how much less procyclical the wage component of the user cost of labor would be if the countercyclicity of these hiring costs is taken into

Figure 2  
**The New-Hire Wage and the User Cost of Labor**



Source: Bils, Kudlyak, and Lins (2023).

Notes: The figure shows the quality-adjusted new-hire wage and the wage component of the user cost of labor with time-varying and history-dependent separation rate and discount factor (left axis). The series are in logs and are normalized to average zero for the sample period (for example, 0.1 means 10 percent above sample-period mean). The unemployment rate (right axis) is in percentage points. All series are annual. Shaded areas mark the NBER-dated recessions.

account. To do this, they adjust the cyclicity of the wage component of user cost of labor to compensate for the “excess” countercyclicity of the nonwage component of the user cost that arises due to the procyclical expected durability of matches.<sup>9</sup> Depending on the assumption about the nonwage costs, the adjusted cyclicity of the price of labor is between  $-4.21$  and  $-4.79$  percent. Comparing these estimates to the one without adjustment, a drop of  $-5.32$  percent, reveals that adjusting for

<sup>9</sup>Bils, Kudlyak, and Lins (2023) consider two different scenarios for the nonwage costs associated with adding a worker. Under the first scenario, a hiring cost is incurred in the starting period and equals to three months of wages. This is relatively large relative to typical values in the literature (for example, as compared to the costs calculated by Silva and Toledo [2013] for hiring and training). Under the second scenario, the nonwage component includes, in addition to the upfront costs, persistent training costs that decline over the match such that the expected discounted value of the flow of the nonwage costs is 0.96 of a full year of steady-state earnings.



match durability reduces the cyclical of the wage component of the user cost by only about one-fifth, even generously calibrating hiring and training costs.

In short, Bils, Kudlyak, and Lins (2023) find that the price of labor is strongly procyclical, driven by a combination of procyclicality of new-hire wages as well as the procyclicality of the effect of locking-in of the future wages in the relationship to the economic conditions at the time of hiring. Out of the total decline of wages of  $-5.32$  percent when the unemployment rate increases by 1 percentage point, 2.3 percentage points reflect a procyclical new-hire wage, while the other 3 percentage points reflect a “lock-in” effect on future wages. Finally, this large procyclicality is somewhat offset by shorter durability of matches that start in recessions.

This estimate of the cyclical of the price of labor is quantitatively similar to other recent studies that estimate the cyclical of the price of labor by its user cost. For example, Kudlyak (2014), Basu and House (2016), Doniger (2021), and Maruyama and Mineyama (2021) find that the price of labor is strongly procyclical, more procyclical than new-hire wage or the average wage.

## **Implications for Volatility of Employment**

The estimated procyclicality of the wage component of the user cost of labor has direct implications for the volatility of employment. It is intuitive to examine the implications in the context of the textbook search and matching model mentioned above. The model’s free entry condition ties labor’s productivity to the sum of the wage component of the user cost of labor and the nonwage component, which in turn is a function of the vacancy-unemployment ratio. As discussed earlier, if the wage component of the user cost of labor is rigid, then any changes in productivity is channeled via the changes in the vacancy-unemployment ratio. In this sense, the free entry condition imposes a trade-off between the elasticity of the wage component of the user cost of labor and the elasticity of the vacancy-unemployment ratio with respect to productivity.

Having estimated the cyclical of the wage component of the user cost of labor, we can check whether the trade-off between the volatility of the wage component of the user cost of labor and of the vacancy-unemployment ratio imposed by the model’s free entry condition holds in the data. It turns out that the restrictions imposed by the model’s free entry condition do not hold in the data (for details, see Kudlyak 2014). Specifically, the cyclical of the wage component of the user cost of labor of above 4 percent translates into the elasticity with respect to productivity of well above 1.5 or even above 2, depending on the sample period. Under the standard parameter values employed in the literature, the model’s free entry condition cannot simultaneously accommodate the high empirical elasticity of the wage component of the user cost and the high empirical elasticity of the vacancy-unemployment ratio. That is, in the data, when productivity declines, both the price of labor and the vacancy-unemployment ratio decline so much that the free entry

condition does not hold. It is as if another force in the model is needed to counter these declines to make the condition hold.

The estimated cyclicalities also has a direct implication for the unemployment volatility puzzle. Specifically, Shimer (2005) points out that in the context of a textbook search and matching model, the observed shifts in productivity are not large enough to generate the observed fluctuations in the vacancy-unemployment ratio. A rigid price of labor is one of the seemingly straightforward solutions. However, the strongly procyclical estimates of the wage component of the user cost of labor imply that not only the data lack rigidity to amplify the volatility of the vacancy-unemployment ratio in the model, but also the price of labor and the vacancy-unemployment ratio are too volatile for the free entry condition in the standard search and matching model to hold. Consequently, the solution to the unemployment volatility puzzle cannot be explained by a wage formation alone, because any wage formation should be able to match the empirical volatility of the wage component of the user cost of labor.

The distinction between the empirical cyclicalities of the wage component of the user cost of labor and of the new hire wage or the average wage is crucial for the conclusion on the propagation of shocks. In the textbook search and matching model of labor markets, wages are typically set by Nash bargaining every period, in new and existing matches. In a model with this method of wage setting, the wage component of the user cost equals the new hire wage, which in turn equals the average wage. In the data, these three wage statistics display vastly different cyclicalities—from strongly procyclicality of the wage component of the user cost of labor to almost rigid average wage. To which of the three empirical measures of cyclicalities does the Nash model's wage component correspond? The empirical counterpart of the wage cyclicalities of a model with Nash bargaining is the cyclicalities of the wage component of the user cost of labor in the data. Nash bargaining each period in all matches serves as a convenient modeling device that delivers a closed-form solution of the wage component of the user cost of labor, but it does not capture history dependence in wages found in the data. Focusing on the cyclicalities of individual wages might lead to a misleading assessment of the quantitative behavior of the model if the wage formation is specified incorrectly.<sup>10</sup> Importantly, the firm's free entry condition imposes the restriction on the wage component of the user cost of labor, not wage per se.

<sup>10</sup>Using the cyclicalities of the average wage instead of the cyclicalities of the wage component of the user cost of labor might lead to an erroneous conclusion that data have the required wage rigidity to solve the unemployment volatility puzzle. For example, in theory, a model with a high flow value of unemployment can generate the empirical volatility of the vacancy-unemployment ratio (Hagedorn and Manovskii 2008). The high flow value of unemployment renders the price of labor in the model to be a large fraction of the output. However, ultimately, the approach rests on the rigidity of the price of labor. Again, the key to whether this approach has support in the data is how rigid the price of labor is. Calibrating the price of labor from the model to the elasticity of the average wages might create an illusion of such rigidity; however, the empirical counterpart of the price of labor is the wage component of the user cost, which is much more procyclical. In fact, there is no value of the unemployment benefits parameter in the set of its feasible values so that the model's free entry condition simultaneously accommodates the

If one disciplines a model by its cyclical price of labor, the corresponding data are the wage components of the user cost of labor. The correct empirical counterpart is especially consequential because in the data a researcher faces a menu of different wage measures—that of an average wage, new hire wage, and the user cost—which range from rigid to strongly procyclical.

## Future Work

To understand the implications of the of the procyclicality of the price of labor, consider, for example, the 2007–2009 recession. Between 2007 and 2009, the unemployment rate went up by 3.5 percentage points relative to its (cubic) trend. According to the estimates of Bils, Kudlyak, and Lins (2023), the estimated price of labor during this time went down by 18.6 percent, a sizeable decline; and more than twice as large as than the decline in the new hire wage. Such a strongly procyclical price of labor suggests that it is relatively cheap to hire in recessions. These results suggest that wage rigidity cannot be the primary cause of cyclical fluctuations in employment. The strong procyclicality of the price of labor suggests that other forces must be at play behind the cyclical fluctuations in labor demand (for an example, see recent work by Kehoe et al. 2023).

With regards to the measurement of the cyclicity of the price of labor, one direction of future research is to examine the cyclicity of the price of labor on the intensive margin. Another promising direction is to examine heterogeneity in the cyclicity of the price of labor across different occupations or, more broadly skill groups. For example, Doniger (2021) estimates more cyclical user cost for college-trained workers compared with those without a college education. Another direction of research involves examining the cyclicity of the price of labor faced by different firms. Different firms face different costs of financing, especially during recessions, which might affect their discount factor and, consequently, the cyclicity of the price of labor that they face. Recognizing the procyclical nature of the price of labor—appropriately understood as the user cost of labor—is reopening fundamental questions about the functioning of labor markets.

■ *I thank Mark Bils, Erik Hurst, and Heidi Williams for insightful comments and suggestions. I thank the managing editor Timothy Taylor for valuable editorial comments and suggestions. The opinions expressed are those of the author and do not reflect those of the Federal Reserve Bank of San Francisco, the Federal Reserve System, or any other organization with which the author is affiliated.*

---

high empirical elasticities of the vacancy-unemployment ratio and of the wage component of the user cost of labor (Kudlyak 2014).

## References

- Barlevy, Gadi.** 2002. "The Sullyng Effect of Recessions." *Review of Economic Studies* 69 (1): 65–96.
- Barro, Robert J.** 1977. "Long-Term Contracting, Sticky Prices, and Monetary Policy." *Journal of Monetary Economics* 3 (3): 305–16.
- Basu, Susanto, and Christopher House.** 2016. "Allocative and Remitted Wages." In *Handbook of Macroeconomics*, Vol. 2, edited by John B. Taylor and Harald Uhlig, 297–354. Amsterdam: North-Holland.
- Baydur, Ismail, and Toshihiko Mukoyama.** 2020. "Job Duration and Match Characteristics over the Business Cycle." *Review of Economic Dynamics* 37: 33–53.
- Beaudry, Paul, and John DiNardo.** 1991. "The Effect of Implicit Contracts on the Movement of Wages over the Business Cycle: Evidence from Micro Data." *Journal of Political Economy* 99 (4): 665–88.
- Bellou, Andriana, and Baris Kaymak.** 2021. "The Cyclical Behavior of Job Quality and Real Wage Growth." *American Economic Review: Insights* 3 (1): 83–96.
- Bils, Mark J.** 1985. "Real Wages over the Business Cycle: Evidence from Panel Data." *Journal of Political Economy* 93 (4): 666–89.
- Bils, Mark, Marianna Kudlyak, and Paulo Lins.** 2023. "The Quality-Adjusted Cyclical Price of Labor." *Journal of Labor Economics* 41 (S1): S13–S59.
- Bowlus, Audra J.** 1995. "Matching Workers and Jobs: Cyclical Fluctuations in Match Quality." *Journal of Labor Economics* 13 (2): 335–50.
- Caballero, Ricardo J., and Mohamad L. Hammour.** 1994. "The Cleansing Effect of Recessions." *American Economic Review* 84 (5): 1350–68.
- Cajner, Tomaz, Leland D. Crane, Ryan A. Decker, John Grigsby, Adrian Hamins-Puertolas, Erik Hurst, and Christopher Kurz.** 2020. "The U.S. Labor Market during the Beginning of the Pandemic Recession." *Brookings Papers on Economic Activity* 51 (S1): 3–33.
- Calvo, Guillermo A.** 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3): 383–98.
- Chodorow-Reich, Gabriel, and Loukas Karabarbounis.** 2016. "The Cyclicalty of the Opportunity Cost of Employment." *Journal of Political Economy* 124 (6): 1563–1618.
- Doniger, Cynthia L.** 2021. "Education and the Margins of Cyclical Adjustment in the Labor Market" Federal Reserve Board Finance and Economics Discussion Paper 2021-019.
- Dunlop, John T.** 1998. "Real and Money Wage Rates." *Journal of Economic Perspectives* 12 (2): 223–34.
- Gertler, Mark, Christopher Huckfeldt, and Antonella Trigari.** 2020. "Unemployment Fluctuations, Match Quality, and the Wage Cyclicity of New Hires." *Review of Economic Studies* 87 (4): 1876–1914.
- Hagedorn, Marcus, and Iourii Manovskii.** 2008. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited." *American Economic Review* 98 (4): 1692–1706.
- Hall, Robert E.** 1980. "Employment Fluctuations and Wage Rigidity." *Brookings Papers on Economic Activity* 11 (1): 91–123.
- Kehoe, Patrick J., Pierlauro Lopez, Virgiliu Midrigan, and Elena Pastorino.** 2023. "Asset Prices and Unemployment Fluctuations: A Resolution of the Unemployment Volatility Puzzle." *Review of Economic Studies* 90 (3): 1304–57.
- Kudlyak, Marianna.** 2014. "The Cyclicalty of the User Cost of Labor." *Journal of Monetary Economics* 68: 53–67.
- Kudlyak, Marianna.** 2024. "Replication data for: How Cyclical Is the User Cost of Labor?" American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E198833V1>.
- Kydland, Finn and Edward C. Prescott.** 1980. "A Competitive Theory of Fluctuations and the Feasibility and Desirability of Stabilization Policy," in *Rational Expectations and Economic Policy*, edited by S. Fischer, 169–98. University of Chicago Press.
- Mankiw, N. Gregory, Julio J. Rotemberg, and Lawrence Summers.** 1985. "Intertemporal Substitution in Macroeconomics." *Quarterly Journal of Economics* 100 (1): 225–51.
- Maruyama, Toshitaka, and Tomohide Mineyama.** 2021. "Cyclicalty and Asymmetry of the User Cost of Labor" Unpublished.
- Mortensen, Dale T., and Christopher A. Pissarides.** 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies* 61 (3): 397–415.
- Mustre-Del-Río, José.** 2019. "Job Duration over the Cycle." *Journal of Money, Credit and Banking* 51 (6): 1691–1711.

- Pissarides, Christopher A.** 2009. "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?" *Econometrica* 77 (5): 1339–69.
- Rotemberg, Julio J., and Michael Woodford.** 1999. "The Cyclical Behavior of Prices and Costs." In *Handbook of Macroeconomics*, Vol. 1B, edited by John B. Taylor and Michael Woodford, 1051–1135. Amsterdam: North-Holland.
- Rudanko, Leena.** 2009. "Labor Market Dynamics under Long Term Wage Contracting." *Journal of Monetary Economics* 56 (2): 170–83.
- Shimer, Robert.** 2004. "The Consequences of Rigid Wages in Search Models." *Journal of the European Economic Association* 2 (2-3): 469–79.
- Shimer, Robert.** 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review* 95 (1): 25–49.
- Silva, José I., and Manuel Toledo.** 2013. "The Unemployment Volatility Puzzle: The Role of Matching Costs Revisited." *Economic Inquiry* 51 (1): 836–43.
- Solon, Gary, Robert Barsky, and Jonathan A. Parker.** 1994. "Measuring the Cyclical Behavior of Real Wages: How Important Is Composition Bias?" *Quarterly Journal of Economics* 109 (1): 1–25.
- Stockman, Alan C.** 1983. "Aggregation Bias and the Cyclical Behavior of Real Wages." Unpublished.
- Thomas, Jonathan, and Tim Worrall.** 1988. "Self-Enforcing Wage Contracts." *Review of Economic Studies* 55 (4): 541–53.
- von Wachter, Till.** 2020. "The Persistent Effects of Initial Labor Market Conditions for Young Adults and Their Sources" *Journal of Economic Perspectives* 34 (4): 168–94.



# Government Data of the People, by the People, for the People: Navigating Citizen Privacy Concerns

Claire McKay Bowen

**A**t 5:00 AM, my Android alarm goes off. When I hit the snooze button, the alarm rings again after ten minutes. To help me wake up, I use the Chrome browser on my phone to access US and New Mexico news articles from sources like the *New York Times* and the *New Mexican*. At 5:30 AM, the Spotify app plays my “Morning Jam” playlist while I head to the local pool. At 6:00 AM, my Garmin watch uses “Indoor Pool Swim” to record my distance, times, speed, and heart rate. When I am done, my watch automatically syncs with the Garmin Connect app, which then updates my other workout apps, TrainingPeaks and Strava. TrainingPeaks allows my triathlon coach to view my workout details, whereas Strava is a social media platform for workout enthusiasts. At 7:35 AM, I use Venmo to pay my swim coach, which records an online payment between two individuals with a note saying, “for June.”

I head home and begin my workday. The Outlook app logs workday start and end times, along with details of meetings and participants. Box and Dropbox record changes made to documents, including additions, removals, and updates. GitHub records “pull requests” when my collaborators and I discuss possible changes to some code. Google Scholar tracks my search history. Overleaf logs access times for working on a paper. Slack records messages in various channels.

When my workday ends, I take my two dogs for a run in an area just outside Santa Fe, New Mexico. Garmin, TrainingPeaks, and Strava record an “Outdoor

■ *Claire McKay Bowen is a Senior Fellow, Urban Institute, Washington, DC. Her email address is [cbowen@urban.org](mailto:cbowen@urban.org).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.2.181>.

Run” with distance, time, speed, and heart rate data. The title on Strava is edited to “EZ Joggo with my Doggos 🐕🐕” for personalization. I capture a photo of our run and share it on Instagram and Facebook, along with pictures of my two cats. I end my day watching the British Bake-off on Netflix, while knitting and using Chart-Minder, a knitting app, to track my progress.

This brief narrative highlights the personal data that private companies collect and record about me on a typical day. Although you do not know other personal details like when and what I eat (although there is an app for that, too), you still get more than a glimpse of my professional and personal life.

Tracking other seemingly innocuous information over time can provide additional disclosure. As one example, I started competing in endurance races, like triathalons, in college. If someone linked my publicly available registration information across multiple race events, they could infer that a change in my last name indicates a marriage or that a new home state implies a move in that particular year.

I give my students an assignment along these lines: For one day, they must document what data are collected about them from the moment they wake up until they go to bed. Based on their data logs, they write an essay addressing whether the data is protected by law, potentially beneficial or harmful uses of the data, and equity and ethical effects of collecting this information.

As my students and I both readily admit, most of us willingly use web browsers, engage with social media platforms, use financial apps, participate in rewards programs, and more. While many of these services are free in monetary terms, these companies use consumer data for profit, such as targeted marketing to specific demographics. As a society, we have (mostly) come to accept this reality, although there are evolving practices and regulations that, to some extent, give people some opportunity to opt out of information sharing.

But the types and amounts of personal information collected keeps expanding. Some universities now require students to install an app that tracks their movements on campus (as reported by Belkin 2020). With this information, professors teaching large classes can know their students’ punctuality, tardiness, or class absences. Beyond attendance records, one can imagine some social good from this data collection. In the event of a natural catastrophe or a mass shooting on campus, the tracking app could alert students, identify safe refuges, and notify emergency contacts. On the other side, this tracking can be invasive, especially for students who rarely leave campus except during holidays, because the university can have a comprehensive record of their locations for 24 hours a day.

Government collects personal data in two main ways. One is through surveys like the Current Population Survey and the American Community Survey that can provide local and state leaders detailed demographic data about individuals and households in the United States. The other is through administrative data collected for purposes of administering programs, like income tax, data collected from employers to run unemployment and workers’ compensation programs, payments made to households or on behalf of households through unemployment



insurance, Medicaid, and others.<sup>1</sup> Again, tracking the individual-level data from these programs over time would reveal more than just looking at a point in time. Moreover, combining government data with proprietary data from private companies that are not subject to federal data consumer privacy laws could reveal more individual data as well.

This essay will focus on the protection of individual-level government data, with an emphasis on survey and administrative data. I like to say that these data are “of the people, by the people, and for the people.” The data are *of the people*, in the sense that people do care about their privacy and their confidential data. Although they may be willing to trade off information a bit at a time to private-sector actors for useful purposes—like my workout and knitting apps—many people would be deeply unhappy if their personal data was widely available.<sup>2</sup> The data are also *by the people*, in the sense that government collection of people’s information is supported by taxpayer dollars. Therefore, one could argue that anonymized individual-level data should be accessible to data users—such as data practitioners, external researchers, or public policymakers. Literally volumes of research could be cited here about how increased access to government data results in social good *for the people*. As one example, Chetty, Friedman, and Rockoff (2014) showed how elementary school teacher quality has a substantial effect on economic outcomes later in life—a finding that was only possible when the economists had direct access to administrative data. For another example, Nagaraj and Tranchero (2023) discovered that applied researchers with confidential data access through the Federal Statistical Research Data Centers are more likely to produce papers that are cited in public policy reports.

While the data privacy community agrees that these data should be more widely accessible, what to protect in that data and how to do so are highly and intensely debated. The community involved in this conflict can be usefully divided into four groups (Williams and Bowen 2023): (1) *data users and practitioners* consume data, such as analysts, researchers, planners, and decision-makers; (2) *data privacy experts or researchers* specialize in developing data privacy and confidentiality methods; (3) *data curators, maintainers, or stewards* are responsible for data safekeeping, and in that sense are sometimes said to “own” the data; and (4) *data intruders, attackers, or adversaries* try to gather sensitive information from the confidential data. The discussion that follows will refer to all four groups.

<sup>1</sup>While “administrative data” often refers to what is collected by the government for programmatic purposes, the term can be more broadly used to cover, for example, other organizational records. For example, colleges and universities have data student applications, class registration, dormitory assignments, and grades, and private companies have administrative data for purposes like tracking orders.

<sup>2</sup>This article uses the terms “data privacy” and “data confidentiality” interchangeably, but readers should note that the terms are sometimes used with separate meanings. In some cases, data confidentiality refers to how the data privacy community protects participants’ information in the data, such as who should have access to the sensitive data under what restrictions. Data privacy refers to the amount of personal information individuals allow others to access about themselves.

Throughout this paper, my discussion will emphasize the fundamental trade-off between data privacy and data usefulness—and how determining an appropriate balance can be difficult. In extreme cases, a perfectly protected dataset would never be released, which is useless to data users. Conversely, a perfectly useful dataset would be released without any statistical data privacy methods applied or any data protections, which would violate privacy concerns. These extreme cases demonstrate how it is impossible to achieve perfect privacy and utility. This is why, even with statistical methods to protect privacy in any given use of data, some information is inevitably leaked with each release of a dataset or statistic. Another challenge is repeated data or statistic releases can gradually erode the overall privacy protection, ultimately reaching a point where the level of protection becomes equivalent to releasing the data without any alterations. This highlights why data curators collaborate with privacy researchers to find a balance between these two extremes and prevent excessive information disclosure.

## **Traditional Methods of Accessing Data**

### **Protecting and Releasing Public Data and Statistics**

For decades, government agencies produced public use data and statistics. During the pre-computer era in the early to mid-1900s, public use data were available to those who braved the “government documents” section of research libraries or those who physically went to government offices to inspect available files. But government agencies typically reported summary statistics, such as total spending on unemployment insurance and total number of people receiving it. Knowing such information on a county-by-county or metro-area basis did not pose much threat to privacy.

As the computer era arrived, government agencies started to provide more detailed public use data that could be directly accessible to both researchers and the general public. For example, the Statistics of Income (SOI) Division of the Internal Revenue Service (IRS) releases a public use file for data users based on administrative taxpayer data. Several organizations, such as the American Enterprise Institute (DeBacker, Evans, and Phillips 2019), the Urban-Brookings Tax Policy Center (McClelland et al. 2019), and the National Bureau of Economic Research (Bierbrauer, Boyer, and Peichl 2021), develop microsimulation models based on this public use file that inform the public on potential impacts of tax policy proposals.

To ensure that these public use data protect individual privacy, extensive statistical data privacy methods are implemented. Here, I will use a fictitious socioeconomic dataset to illustrate a range of such methods. For a more comprehensive overview of these methods, Matthews and Harel (2011) offer a detailed review, while McKenna and Haubach (2019) summarize the specific statistical data privacy methods employed by the US Census Bureau.

Suppose the fictitious micro-level socioeconomic dataset contains hundreds of records for individuals residing in Santa Fe. The top panel of Table 1 displays a

Table 1A

**Fictitious Santa Fe, New Mexico, Socioeconomic Data**

*A fictitious socioeconomic dataset with participants' names, ages, education levels, and income*

<i>Name</i>	<i>Age</i>	<i>Education</i>	<i>Income</i>
Peter	63	Master's	\$51,214
Patricia	48	Bachelor's	\$89,464
Ryan	24	Bachelor's	\$27,893
Rachel	58	Bachelor's	\$74,770
Steve	17	High school	\$623
Suzanne	32	Doctorate	\$135,883
Tomas	81	Bachelor's	\$0
Tiffany	21	Some college	\$11,428

*Source:* Author.

Table 1B

**Adjusting Data to Preserve Anonymity**

*The fictitious socioeconomic dataset (Table 1A) that has been altered with statistical data privacy methods*

<i>ID</i>	<i>Age</i>	<i>Education</i>	<i>Income</i>
01	60	Graduate degree	\$51,000
02	55	Bachelor's	\$89,000
03	25	Bachelor's	\$27,900
04	52	Bachelor's	\$75,000
05	19	No college	\$620
06	37	Graduate degree	\$136,000
07	85	Bachelor's	\$0
08	17	No college	\$11,400

*Source:* Author.

sample of eight records from the dataset, which includes the person's name, age, education, and income. The bottom panel shows how these data have been adjusted to preserve anonymity, using a series of steps.

As a first step, most personally identifiable information, such as names, should be removed from the data. An obvious step is to replace names with numbers. Data curators, who are responsible for safeguarding the data, may generate individual-level identification numbers if they plan to link the data with other information. If there is no intention to link the data with another source, the variable may be entirely removed.

After removing the personally identifiable information, the most common statistical data privacy method is suppression, which involves the removal of certain values from the data. This approach is easy and quick to implement. As an example, when I attended high school in a remote area of Idaho, I was the only Asian American student. Even with names removed, a data intruder could identify me in a

dataset that included information on race/ethnicity. To ensure my privacy, such information could be removed or suppressed.

Another privacy concern in the fictitious dataset is the reporting of income values to the nearest dollar. To make the records less identifiable, we can round the income values. Instead of rounding to the nearest hundred or thousand, some rounding methods introduce randomization in rounding up or rounding down significant figures. For instance, consider an individual with an income of \$596. If we want to round the value to the closest \$10, then there is a 60 percent probability of rounding the income up to \$600 and a 40 percent probability of rounding it down to \$590. There are also other rounding schemes, such as the one utilized by the US Census Bureau, which we implement for the fictitious dataset in the bottom panel of Table 1. In this approach, \$0 is rounded to \$0, \$1–7 rounded to \$4, \$8–\$999 rounded to nearest \$10, \$1,000–\$49,999 rounded to nearest \$100, and \$50,000+ rounded to nearest \$1,000.

Another statistical data privacy method is known as generalization, aggregation, or categorical thresholding. When applying this method, the detailed information is consolidated into broader categories. In our example, we can generalize the education groups, which would decrease or eliminate the number of distinct observations. The bottom panel of Table 1 demonstrates how we changed the education levels of “high school,” “some college,” “bachelor’s,” “master’s,” and “doctorate” into broader categories such as “no college,” “bachelor’s,” and “graduate degree.”

Adding or subtracting random values is another popular statistical data privacy method. One way to generate random values is within specific boundaries (for example, –10 to 10) or based on a probability distribution (for example, a bell curve centered at zero). This method is known as adding noise, injecting noise, sanitizing results, or perturbing the data. In the bottom panel of Table 1, noise has been added to the age variable, resulting in new age values. The random noise is drawn from a bell curve–shaped distribution, such as a normal or Gaussian distribution. We see that some of the added or subtracted values are very small (like 0, 1, and 2), while a few are larger values (for example, 6 and 7). Introducing random values creates some uncertainty, making it more challenging for a malicious actor to discern the original age value.

### **Accessing Federal Data Directly**

Over the years, government agencies have been moving slowly toward allowing more data users direct access to the underlying cleaned data,<sup>3</sup> under strict controls. An example of direct data access is through a secure enclave, such as the Federal Statistical Research Data Centers.<sup>4</sup> This secure enclave became available in 1982

<sup>3</sup>Privacy researchers often distinguish two types of confidential data: the original data and confidential data (Hu and Bowen 2023). The former is the uncleaned, unprotected version of the data, such as the raw census microdata. The latter is cleaned—edited for inaccuracies or inconsistencies—and stripped of some personally identifiable information like names. This essay focuses on the latter.

<sup>4</sup>Federal Statistical Research Data Centers are partnerships between federal statistical agencies and leading research institutions. FSRDCs provide secure environments supporting qualified researchers

(then called the Center for Economic Studies), after data users demanded access to better quality data when the US Census Bureau became more aggressive with its applications of statistical data privacy methods on its data products.

Although more secure facilities are becoming available (for example, the National Science Foundation Secure Data Access Facility<sup>5</sup>), researchers face several challenges to obtain this direct access. Full access to these data is only available to select government agencies, a limited number of data users working in collaboration with analysts from those agencies, or through highly selective research programs administered by these agencies. Further, data users are often required to be US citizens, undergo lengthy clearance processes to gain direct access (which can take months or years), and submit extensive research proposals.

Another challenge is accessibility to these secure facilities. The 33 Federal Statistical Research Data Centers across the United States may seem like enough to be geographically accessible to most data users. But that is not the case. These data centers are primarily located in places with large academic institutions. For me, living in Santa Fe, New Mexico (the state capital), the closest is a 7.5-hour drive to Boulder, Colorado. Moreover, remote access to these data centers grants access to only a limited selection of confidential data and requires a setup that simulates a secure enclave working station, which may prove challenging for many individuals.

## **Developing Another Tier of Data Access: Verification and Validation Servers**

In recent years, synthetic data generation has become a popular method for producing public data that releases more useful information than the other traditional statistical data privacy methods while protecting privacy. The general concept of synthetic data is generating pseudo or fake records that preserve the structure and statistical relationships in the confidential data. Ever since Rubin (1987) proposed the original method to create multiple imputations for missing data, the research community has developed several different flavors (partially and fully synthetic) and approaches to generate synthetic data (such as Bayesian synthesis models). Hu and Bowen (2023) provide a more detailed review of synthetic data for privacy protection.

There are obvious concerns over whether a new analysis carried out on synthesized data will reflect the underlying data. As an example, the Urban Institute and Statistics of Income Division at the IRS have created a synthetic version of the SOI public use file that matches certain statistical properties of the underlying data, such

---

using restricted-access data while protecting respondent confidentiality.” For details, see <https://www.census.gov/about/adrm/fsrdc.html> (accessed on January 15, 2024).

<sup>5</sup>The National Science Foundation Secure Access Facility provides authorized researchers secure remote access to National Center for Science and Engineering Statistics data and metadata, such as the Survey of Earned Doctorates and the national Survey of Recent College Graduates (<https://www.norc.org/research/projects/nsf-secure-data-access-facility.html>, accessed on January 15, 2024).

as means, variances, and covariances, and provides reliable estimates from micro-simulation models (Bowen et al. 2022a, 2022b). The drawback is that this synthetic SOI public use file (and any other synthetic data) may not perform well for other types of analyses that were not accounted for when synthesizing the data, and thus may not yield reliable parameter estimates for more complex statistical models.

To address this issue, the data privacy community has proposed the use of verification and validation servers, which we delineate here following Williams et al. (2023). A “verification server” is a system that provides information about the quality of statistical inference derived from publicly released data. In other words, the verification server might report whether inferences about the sign, statistical significance, magnitude of the estimates, or other elements derived from the public data or synthetic data are consistent with the confidential data. As one example, Barrientos et al. (2018) synthesized data and created a pilot verification server for the US Office of Personnel Management, which allows the study of career paths and pay differentials for federal employees.

A “validation server” allows users to submit and run statistical analyses on the confidential data after the users have developed those analyses using the publicly released data. For instance, data users could apply a tax microsimulation model on the synthetic public use file. If the data users have other statistical analyses, they can then develop and debug programs using the synthetic data to ensure those programs will run seamlessly on the validation server, so long as synthetic data record layouts are identical to the confidential data, and receive statistically valid results. For instance, until September 2022, the US Census Bureau used to support validation servers for two experimental synthetic databases via the Synthetic Data Server at Cornell University: the Synthetic Longitudinal Business Database (Kinney, Reiter, and Miranda 2014)<sup>6</sup> and the Survey of Income and Program Participation’s (SIPP) Synthetic Beta Data Product (Benedetto, Stanley, and Totty 2018).<sup>7</sup>

The idea of tiered access has been proposed in several public policy discussions. For instance, the Foundations for Evidence-Based Policymaking Act of 2018 (often called the “Evidence Act”) “requires [federal] agency data to be accessible and requires agencies to plan to develop statistical evidence to support policy-making” (see <https://www.congress.gov/bill/115th-congress/house-bill/4174>). The Evidence Act also calls for the establishment of a National Secure Data Service, which could serve as a host for data, validation servers, and verification servers. In 2022, the Advisory Committee on Data for Evidence Building, a congressional committee charged with “reviewing, analyzing, and making recommendations on how to promote the use of Federal data for evidence building,” released a final

<sup>6</sup>See the “Synthetic Longitudinal Business Database (SynLBD),” at <https://www.census.gov/programs-surveys/ces/data/public-use-data/synthetic-longitudinal-business-database.html> (accessed on January 15, 2024).

<sup>7</sup>See the “Synthetic SIPP Data,” at <https://www.census.gov/programs-surveys/sipp/guidance/sipp-synthetic-beta-data-product.html> (accessed on January 15, 2024).

report recommending the creation of tiered access systems.<sup>8</sup> During the summer of 2023, America's Datahub Consortium, a collaboration sponsored by the National Center for Science and Engineering Statistics within the National Science Foundation, posted nine requests for solutions. Several of these requests involve creating demonstration products to support a tiered access model and explore the use of verification metrics in validating estimates produced from public data.<sup>9</sup>

## **Implementing Formal Privacy for Tiered Data Access**

Although the Synthetic Data Server at Cornell University provided access to confidential data, it was not automated. The demand for the service often exceeded available staff time, causing long delays for approval. Another drawback is that the manual review processes involve high-level staff reviewing program code and output to identify potential disclosures of confidential data. This manual review process is time-consuming, based on human subjectivity, and imperfect. Also, because the staff reviewed proposals one at a time, they did not appropriately account for the cumulative disclosure risk over time.

A well-designed and automated validation server system could provide consistent and robust privacy protection with little or no human review, which is both safer and less labor-intensive than past manual review of research programs that involve subjective human review. To complement the synthetic SOI public use file, the Urban Institute and Statistics of Income Division at the IRS are developing an automated validation server that uses differentially/formally private methods to release statistically valid results with privacy protections (Barrentios et al. 2023; Taylor et al. 2021). Before delving into the practical challenges of such a system, I will provide a general background on differential/formal privacy.

### **Making a Noisy Case for New Privacy Definitions**

The illustrative socioeconomic dataset provided above showed how implementing statistical data privacy methods to release public data and statistics is a viable privacy-protecting alternative for data users who are unable to access confidential data directly. Yet, like direct data access, these traditional approaches have their limitations. One challenge is predicting the behavior of data intruders, making it difficult to determine what information should be considered sensitive. A notorious example of how a malicious actor could gain private information from seemingly anonymous data is the Netflix Prize. Based on the movies that Netflix users liked in the past, Netflix wanted to recommend future movies that people will

<sup>8</sup>See the "Advisory Committee on Data for Evidence Building (ACDEB) releases final report (10.14.22)," at <https://www.aeaweb.org/forum/3174/advisory-committee-evidence-building-acdeb-releases-report> (accessed on July 11, 2023).

<sup>9</sup>See "Opportunities and Pending Awards," at <https://www.americasdatahub.org/opportunities/> (accessed on July 11, 2023).

rate highly. Thus, back in 2006, Netflix organized a competition with a \$1 million prize for an algorithm that would look at past movie ratings and then lead to users ranking the recommended movies more highly—specifically, developing a recommendation system to beat the existing Netflix algorithm by 10 percent. For the contest, Netflix provided a dataset comprising of about 100 million movie ratings from nearly 500,000 anonymous users.

However, instead of using the dataset to improve movie rating predictions, Narayanan and Shmatikov (2008) showed that they could reidentify the supposedly anonymized records in the dataset. They achieved this by using publicly available information from ratings on the Internet Movie Database (IMDb), an Amazon.com owned online database of actors, movies, TV series, and video games. Although knowing a person's movie rating may appear innocent, a data attacker could use preferences about movies to infer more sensitive aspects of people's lives, such as their sexual orientation or political preferences. Netflix did award the \$1 million prize, but faced with lawsuits over privacy concerns, the company cancelled its planned follow-up contest in 2010 (as reported in Lohr 2010).

In a more recent example, the *New York Times* acquired a large cell-phone tracking dataset that contains only time and location information. In the paper, Thompson and Warzel (2019) wrote an article about how they successfully identified an individual based on deviations from their usual work routine, specifically noting a shift from commuting to Microsoft to commuting to Amazon.

These examples illustrate how traditional methods of protecting privacy can fall short. Data curators and privacy experts often struggle to predict what resources a data intruder has at their disposal. Does a data intruder have access to other databases or considerable computing power? How could a data intruder use these resources to exploit a combination of public data and other publicly available statistics? Without predicting the data intruder's behavior, data curators and privacy experts will have difficulties accounting the cumulative disclosure risk from a series of data releases, occurring each time a public dataset or statistic is released based on confidential data, which might gradually lead to revealing personal information.

Another drawback of traditional statistical data privacy methods is their lack of transparency, which hinders reproducibility and replicability. The traditional methods of protecting individual privacy often rely on the concept of "security through obscurity," where parts of these methods are concealed to prevent clever data intruders from reverse engineering and recreating the confidential data. However, as various professional and research organizations advocate for reproducibility and replicability in research,<sup>10</sup> traditional statistical data privacy methods will fall short on this dimension too.

<sup>10</sup>Examples include the American Economic Association (in its "Data and Code Availability Policy," at <https://www.aeaweb.org/journals/data/data-code-policy>) and the American Statistical Association (in its "ASA Journal Policies on Data Sharing and Reproducibility," at <https://www.amstat.org/publications/q-and-as/asa-journal-policies-on-data-sharing-and-reproducibility>).



### Introducing Formal Privacy

The relatively new mathematical concept of differential privacy addresses the challenges that traditional statistical data privacy methods face, revolutionizing the field of data privacy and confidentiality. A report done for the US Census Bureau defined these privacy definitions, or “formal privacy” definitions, as a subset of statistical data privacy methods that provide “formal and quantifiable guarantees on inference disclosure risk and known algorithmic mechanisms for releasing data that satisfy these guarantees” (JASON 2022, p. 41). In other words, what makes a privacy definition formally private is broadly the ability to (1) quantify and adjust the privacy-utility trade-off (typically through parameters); (2) rigorously and mathematically prove the maximum privacy loss that can result from the release of information; and (3) compute total disclosure risk or privacy loss from multiple individual information releases (Bowen and Garfinkel 2021). I will refer to differential privacy and related privacy definitions that satisfies these characteristics as “formal privacy.”<sup>11</sup>

Although noise addition has existed before formally private methods (as in the example earlier in the paper), the ability to account for the total privacy loss is crucial in ensuring strong privacy protection. Formal privacy methods are akin to using a debit card with a predetermined budget, whereas traditional statistical data privacy methods are like a limitless credit card. In both cases, there is a cumulative cost of associated purchases, but only the debit card requires constant monitoring of the remaining balance. In both traditional and formal privacy settings, data curators must restrict the type and quantity of queries made to the data. However, a formally private framework requires data curators to exercise diligence in tracking the usage of data to ensure privacy protections.<sup>12</sup>

Since its inception, differential privacy has been the most well-known formal privacy definition (Dwork et al. 2006). As mentioned earlier, differential privacy adheres to strict mathematical conditions to be considered differentially private, which is not a statement or description regarding data itself (in other words, differentially private methods are data agnostic). Thus, differential privacy does not make assumptions about what a data intruder considers sensitive information, nor what external data or computational power the intruder has access to, now or in the future. Differential privacy instead assumes the worst-case scenario that the data intruder has information on every record in the confidential data except one, unlimited computational power, and the record that the intruder has no information on is the most extreme possible record (or an extreme outlier) that could alter the target statistic or information that a data curator wants to release publicly.

This worst-case scenario assumption also enables data curators to disclose details of differentially private and formally private methods. For instance, privacy

<sup>11</sup> A factor that led to the generalized term of formal privacy is that privacy experts developed alternative versions of differential privacy that “relaxed” differential privacy’s strong privacy guarantees (that is, the worst-case scenario conditions listed earlier). Although differential privacy is a more common definition, most practical applications use a relaxation or an alternative formally private definition, such as the 2020 Census (Bowen, Williams, and Pickens 2022).

<sup>12</sup> This analogy is borrowed from Bowen and Garfinkel (2021).

researchers can publish the privacy parameter values often referred to as privacy-loss budgets. This openness about methods contrasts with traditional statistical data privacy methods, such as when the US Census Bureau implemented “data swapping,” a statistical data privacy method that involves exchanging observations with similar variable characteristics, from the 1990 Census to the 2010 Census. Throughout that time, the Census Bureau did not report the swapping rate, due to the risk of a data intruder reverse-engineering the method.

### **Purchasing Statistics with a Privacy-Loss Budget**

The other key feature of formal privacy definitions is their ability to account for the cumulative disclosure risks or privacy-loss with the public release of information derived from confidential data. These definitions use the concept of a privacy-loss budget that is typically represented as  $\epsilon$ . (There are additional privacy parameters for various definitions, such as  $\delta$  and  $\rho$ . For simplicity, I will refer to the privacy budget as  $\epsilon$ .) The data curator can treat the privacy-loss budget or privacy parameter like a knob to adjust the trade-off between privacy and utility when releasing a statistic of dataset publicly. This means the data curator must set the privacy-loss budget before publishing any information publicly and must track the budget or risk exhausting the budget prematurely. Similar to a financial budget, if the privacy-loss budget is exhausted, then no more information from the confidential data would be released.

Earlier in the article, I stated how a perfectly protected dataset is one that is never released, and a perfectly useful dataset is one that is released unaltered. Formal privacy frameworks can explain these extreme scenarios with  $\epsilon$ . When  $\epsilon$  becomes very large (approaching infinity), the released dataset is unaltered, but has no privacy. When  $\epsilon$  becomes very small (approaching zero), the released dataset has maximum privacy, but no utility. Finding the balance between the two extremes is easier with the privacy-loss budget, because the data curator can increase the privacy-loss budget if they desire a more accurate statistic (but less privacy) or want more privacy (but less accuracy).

Data curators can also allocate the privacy-loss budget over several public datasets and statistics. As an example, imagine the privacy-loss budget is a monthly budget for household expenses (like spending on housing, groceries, utilities, and transportation). Some people might want to allocate their budget equally to each category of monthly expense; others might want to allocate more of their monthly budget to groceries than transportation. Similarly, some data curators might prioritize releasing multiple statistics, while other data curators might allocate the full privacy budget to allow the release of a more detailed microdata. The privacy-loss budget empowers data curators on how they can allocate and account for each individual release of information, while maintaining the overall budget for the system.<sup>13</sup>

<sup>13</sup>Analogy is borrowed from Bowen, Williams, and Pickens (2022).

### **Calculating the Global Sensitivity of Statistics**

Most formally private methods use the concept of “global sensitivity,” which describes how resistant a statistic is to the presence of outliers, tying into the “spending power” of the privacy-loss budget. If a statistic is more robust or resistant to outliers, less privacy-loss budget is needed to release a more accurate statistic. For the converse, a statistic that is less robust or resistant to outliers will require more privacy-loss budget to release a more accurate statistic.

The concept of global sensitivity works by quantifying how much an output can change with the addition or removal of the most extreme possible record that could possibly exist in the population. This means that regardless of whether that record is present in the data, we must consider that the record could be in the data—this is why formally private methods are data agnostic. Simply put, formally private methods use global sensitivity to account for any possible version of the data that could exist, protecting both against future data releases and new technologies.

Imagine the data that need protection contains socioeconomic information, and the question being asked is, “What is median wealth?” Within a formally private framework, it must consider the most extreme possible record that could exist in any given data that have demographic and financial information. In this example, that person is Elon Musk, who was the wealthiest person in the world at the end of 2023 according to *Forbes* magazine.<sup>14</sup> If Musk is present or absent in the data, the median wealth should not change much. This means a more accurate answer could be provided with fewer alterations to the median income statistic, because the statistic is less sensitive to the extreme outlier, Musk. In contrast, consider the question: “What is the average wealth?” Unlike the previous statistic, the answer to that statistic would drastically change if Musk were present or absent from the data. To protect the extreme case at a given level of privacy-loss budget, a formally private algorithm would need to provide a significantly less accurate answer by altering the statistic more.<sup>15</sup>

### **Highlighting Practical Challenges of an Automated Validation Server with Formal Privacy**

With the basics of formal privacy in place, we can imagine how an automated validation server could enable more data users to have access to the underlying administrative data, without placing additional burdens on government staff. At a high level, an automated validation server using formal privacy would require the data curator to monitor a privacy budget, allowing the curator to track the cumulative effect of several submitted analyses from data users.

While the concept of formal privacy holds promise for automated validation and verification servers, it poses several implementation challenges. To help explain these challenges, I will discuss an example: a collaboration between the

<sup>14</sup>For details, see “The World’s Real Time Billionaires,” available at <https://www.forbes.com/real-time-billionaires> (accessed on January 15, 2024).

<sup>15</sup>Analogy is borrowed from Bowen, Williams, and Pickens (2022).

Urban Institute and Statistics of Income Division at the IRS to create an automated validation server applying formally private methods that enables safe access to confidential administrative tax data.

One challenge has been the discovery that the formal privacy field is not as far along as it may have seemed a few years ago. Barrientos et al. (2023) conducted a feasibility study on the state-of-the-art formally private methods to release summary statistics and regression coefficients, with evaluations on administrative tax data from the Statistics of Income Division at the IRS and survey data from the Census Bureau. They found that formally private methods for summary statistics perform well for small privacy-loss budgets, which means that for this purpose, formal privacy provides more “bang for your buck.” In contrast, the formally private regression methods require much larger privacy-loss budgets. Additionally, very few of these formally private methods provided standard errors for regression coefficients, which are essential for most social science research. The authors highlighted that more research is needed to develop methods that are robust to data types that do not involve normal distributions, and to calculate the uncertainty around the estimates. There is a lot of research on how to implement formally private methods for prediction or the outcomes of regression models.

Based on research studies and discussions with privacy experts, data users, and the Urban Institute project team, Snoke et al. (2024) identified five types of incompatibilities between current practices in statistical data analysis and data privacy—specifically, estimates for traditional statistical inference, control or nuisance variables, assumptions on the range of the data or other assumptions, performing exploratory data analysis, and limited queries and the privacy budget. I will discuss the last two.

Most data users need to perform exploratory data analysis and subsequent analyses. In many cases, data users may be unsure of the specific analyses they want to conduct until they have access to the data. Because data users will have a limited privacy budget, a validation server can provide a disincentive for undesirable research practices like p-hacking (that is, searching through multiple, alternative specifications to find one that meets the criterion for statistical significance). However, concerns remain regarding how a formally private validation server can handle other research practices, such as multiple testing or when a journal reviewer requests that alternative models be applied.

This challenge sheds light on a broader issue for formal privacy: setting privacy budgets for data users. How should data curators allocate privacy budgets to different data users? How can data users determine their model specifications without depleting their entire budget? How do data users conduct robustness checks without exhausting their allocated resources? A related issue arises when multiple data users submit the same analysis. Suppose data user A queries a statistic, and two months later, data user B queries the same statistic on the same part of the data. How should the validation server handle this scenario? There are two general options: (1) treat queries A and B as separate queries (with two different results) and charge both users their respective privacy-loss budgets (which could be the

same or different amounts); or (2) use the privacy-preserving result from A for B. Both have pros and cons.

The former approach avoids the conflict of notifying data user B that data user A has already conducted the analyses. After all, researchers may not want others to know what kind of analyses they are conducting for various reasons, such as to avoid being scooped. But submitting the queries separately would result in data user A and data user B having slightly different answers to the same statistic. This creates communication and education problems in explaining to both data users that their answers are valid (a problem to be discussed further in the next section).

The latter approach would avoid the confusion of having two different answers. The data users could also split the privacy-loss, reducing the cost to their respective privacy-loss budgets. A drawback is that both data users would be informed that another has conducted a similar analysis. An additional complication is if data user B wants a more accurate result and, therefore, wants to spend a higher privacy-loss budget than data user A, who does not want to exceed a certain privacy-loss amount.

### **Measuring the Privacy-Loss Budget**

Although the concept of privacy-loss budgets addresses the ad hoc nature of traditional statistical privacy methods, it brings other complicated issues. The incompatibility problems discussed earlier are just some of many others needing to be addressed, such as most statistical data privacy methods require the data user to make assumptions without access to the data (Snok et al. 2024).

The challenge of bringing formal privacy directly to real-world datasets is an ongoing subject of applied research. As an example of a central problem, one of my colleagues working at the Urban Institute asked, “Does it make sense to think of a unit of privacy-loss budget as having the same value across statistics, functions, and runs?”

One might assume that a given unit of privacy-loss should be considered equally valuable across different statistical analyses, but, in reality, this seems unlikely. As a metaphor, suppose you have \$20. Regardless of where you are in the United States, that \$20 holds the same face value. However, in one area, it might be sufficient to buy a decent meal at a specific restaurant, while in another area, it may not be enough. Furthermore, what defines a good meal may differ across people. Similarly, a certain quantity of privacy may seem satisfactory in some contexts, but not in others, and it may satisfy some people, but not others.

When I shared this analogy with another colleague, the response was, “I like the metaphor that you don’t always get what you pay for. But, to me, the currency is more like Martian bucks instead of US dollars.” My colleague’s point highlights the fact that the field of data privacy and confidentiality has no inherent understanding of the actual value of the privacy-loss budget currency. All we know is that increasing the privacy-loss budget should result in more accurate analysis or produce a higher quality dataset. Because we do not know what a privacy-loss budget truly affords us, it might as well be from another world.

## A New Landscape of Data Privacy

Data curators, stakeholders, privacy experts, and data users all face challenges concerning a desire to expand the use of government data, especially in building linkages between administrative and survey data. Unless and until data privacy is securely protected, such data are likely to be available only narrowly, or not at all.

### Educating the Data User Community

Little is known about the expectations and needs of data users in general—let alone their understanding and perceptions of formal privacy. Williams et al. (2023) conducted a convenience sample survey of economists from the American Economic Association on their baseline knowledge about differential/formal privacy, attitudes toward differentially/formally private frameworks, types of statistical methods that are most useful to economists, and how the injection of noise under formal privacy would affect the value of the queries to the user. At a high level, the survey found that most economists are unfamiliar with formal privacy and differential privacy (and if they know about it, they are skeptical). Instead, economists rely on simple methods for analyzing cross-sectional administrative data but have a growing need to conduct more sophisticated research designs, and economists have low tolerance for errors, which is incompatible with existing formal privacy definitions and methods.

The results from the Williams et al. (2023) survey are not surprising. In general, traditional statistical data privacy methods are more intuitive and easier to explain, such as why data curators should remove unique records. In contrast, formally private methods are more complex and lack an intuitive definition.

Although there has been an explosion of new communication materials to explain formal privacy and other data privacy concepts,<sup>16</sup> such efforts are trying to fill a chasm and we are not even close. To put it into perspective, if we asked random economists to recommend their favorite education or communication materials about, say, machine learning or artificial intelligence, many would have a favorite book or blog series in mind. They may even have suggested materials that are more focused on concepts, or a perspective from a certain field, or on coding. If we asked random economists the same question, but for data privacy in the context of safe access to administrative and survey data, they likely would have few recommendations or even none at all.

One way to address the lack of education and communication materials is to teach the next generation and increase the number of those in the field. Yet despite the need for data privacy education, most higher education institutions do not offer dedicated courses on the topic. When data privacy is taught, it is typically at the graduate level within computer science departments. Some undergraduate professors who research data privacy and confidentiality may introduce these topics in

<sup>16</sup>One of my favorites is a video created by minutephysics for the US Census Bureau, available at “Protecting Privacy with MATH (Collab with the Census),” <https://www.youtube.com/watch?v=pT19VwBAqKA> (accessed on June 21, 2023).

seminar courses, but they are not usually stand-alone courses. As a result, individuals with technical backgrounds outside of computer science, such as economists, are greatly underrepresented in this important area of study. Therefore, other departments outside of computer science should consider hosting their own statistical data privacy courses or incorporating these concepts into existing courses. When integrating these statistical data privacy concepts, professors can encourage students to consider the legal, social, and ethical implications of data privacy, ethics, and equity. They can also delve into the principles of data guardianship, custodianship, and data permissions (Williams and Bowen 2023).

### **Addressing Data Equity in Data Privacy**

The methods used to protect individuals' information do not always have an equal impact on all groups represented in the data. A published dataset might ensure the privacy of people who are the majority in the dataset but fail to ensure the privacy of those in smaller groups. Similarly, alterations to the data may be more useful for learning about some groups more than others. Ultimately, how entities collect and share data can have varying effects on underrepresented groups of people.

Although there are many discussions on data equity *and* data privacy, few conversations focus on equity *in* privacy. In light of this, Bowen and Snoke (2023) developed a guide as part of the “Do No Harm Guides” series. This fourth installment of the series focuses on exploring the current state of equity-focused work in statistical data privacy. The authors conducted interviews with nine experts in privacy-preserving methods and data-sharing, including researchers and practitioners from academia, government, and industry sectors with diverse technical backgrounds. The authors asked about the experience of these experts in implementing statistical data privacy methods and how they define equity in the context of privacy, among other topics. The authors then created an illustrative example to highlight potential disparities that can result from applying various statistical data privacy concepts (including suppression, synthetic data, and differential privacy) without an equitable workflow. Here are some of their key takeaways: do not treat equity as a separate field of study; work with groups represented in your data; and there is no methodological silver bullet.

### **Engaging with Data Privacy Issues**

There are a few prominent options for learning about data privacy methods and becoming involved in these topics besides becoming a privacy researcher. For instance, the Joint Program in Survey Methodology at the University of Maryland has been offering a course on synthetic data.<sup>17</sup> The Urban Institute offered an all-day course at the 2023 Joint Statistical Meetings, where the instructors introduced the

<sup>17</sup>See “Synthetic Data: Balancing Confidentiality and Quality in Public Use Files,” a course by Joerg Drechsler and Jerome P. Reiter, at <https://jpsm.umd.edu/academics/jpsm-short-course-2023-24-schedule> (accessed on June 21, 2023).

basics of data privacy.<sup>18</sup> The Urban Institute has also offered similar trainings for the Bureau of Economic Analysis, Allegheny County, and the Statistics of Income Division.

There is currently no dedicated conference focused on the intersection of data privacy and public policy, but interest in the field is growing. In 2023, the National Bureau of Economic Research<sup>19</sup> and National Institute of Statistical Sciences<sup>20</sup> hosted separate data privacy workshops that brought together privacy experts and data users. Attendees from these workshops are currently organizing the first ever Privacy and Public Policy Conference in 2024 with the goal “to foster and enhance collaboration among privacy experts, researchers, data stewards, data practitioners, and public policymakers.”<sup>21</sup> With the recent surge of venues, the time is obviously ripe to help shape the future of data privacy, make meaningful contributions to its policy debates, and ensure the responsible representation of people in data.

<sup>18</sup> See “Introduction to Data Privacy and Data Synthesis Techniques,” a course by Aaron R. Williams and Claire McKay Bowen, at <https://ww2.aievolution.com/JSMAnnual/index.cfm?do=ev.pubSearchEvents> (accessed on June 21, 2023).

<sup>19</sup> See “Data Privacy Protection and the Conduct of Applied Research: Methods, Approaches, and their Consequences, Spring 2023,” hosted by the National Bureau of Economic Research, at <https://www.nber.org/conferences/data-privacy-protection-and-conduct-applied-research-methods-approaches-and-their-consequences> (accessed on June 21, 2023).

<sup>20</sup> See “IOF Workshop: Advancing Demographic Equity with Privacy Preserving Methodologies,” hosted by the National Institute of Statistical Sciences, at <https://www.niss.org/events/iof-workshop-advancing-demographic-equity-privacy-preserving-methodologies> (accessed on June 21, 2023).

<sup>21</sup> See “Privacy and Public Policy Conference,” at <https://privacypublicpolicy-conference.github.io/website/> (accessed on January 15, 2024).



## References

- Barrientos, Andrés F., Alexander Bolton, Tom Balmat, Jerome P. Reiter, John M. de Figueiredo, Ashwin Machanavajjhala, Yan Chen, Charley Kneifel, and Mark DeLong. 2018. "Providing Access to Confidential Research Data through Synthesis and Verification: An Application to Data on Employees of the US Federal Government." *Annals of Applied Statistics* 12 (2): 1124–56.
- Barrientos, Andrés F., Aaron R. Williams, Joshua Snoke, and Claire McKay Bowen. 2023. "A Feasibility Study of Differentially Private Summary Statistics and Regression Analyses with Evaluations on Administrative and Survey Data." *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.2023.2270795>.
- Belkin, Douglas. 2020. "No Place to Hide: Colleges Track Students, Everywhere." *Wall Street Journal*, March 5. <https://www.wsj.com/articles/the-many-ways-college-students-may-be-tracked-on-campus-11583354852>.
- Benedetto, Gary, Jordan C. Stanley, and Evan Totty. 2018. "The Creation and Use of the SIPP Synthetic Beta v7.0." US Census Bureau Working Paper.
- Bierbrauer, Felix J., Pierre C. Boyer, and Andreas Peichl. 2021. "Politically Feasible Reforms of Nonlinear Tax Systems." *American Economic Review* 111 (1): 153–91.
- Bowen, Claire McKay, Victoria Bryant, Leonard Burman, John Czajka, Surachai Khitatrakun, Graham MacDonald, Robert McClelland et al. 2022a. "Synthetic Individual Income Tax Data: Methodology, Utility, and Privacy Implications." In *Privacy in Statistical Databases*, edited by Josep Domingo-Ferrer and Maryline Laurent, 191–204. Cham: Springer Nature.
- Bowen, Claire McKay, Victoria Bryant, Leonard Burman, Surachai Khitatrakun, Robert McClelland, Livia Mucciolo, Madeline Pickens, and Aaron R. Williams. 2022b. "Synthetic Individual Income Tax Data: Promises and Challenges." *National Tax Journal* 75 (4): 767–90.
- Bowen, Claire McKay, and Simson Garfinkel. 2021. "Philosophy of Differential Privacy." *Notices of the American Mathematical Society* 68 (10): 1727–39.
- Bowen, Claire McKay, and Joshua Snoke. 2023. *Do No Harm Guide: Applying Equity Awareness in Data Privacy Methods*. Washington, DC: Urban Institute.
- Bowen, Claire McKay, Aaron R. Williams, and Madeline Pickens. 2022. "Decennial Disclosure." Washington, DC: Urban Institute.
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–79.
- DeBacker, Jason, Richard W. Evans, and Kerk L. Phillips. 2019. "Integrating Microsimulation Models of Tax Policy into a DGE Macroeconomic Model." *Public Finance Review* 47 (2): 207–75.
- Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. "Calibrating Noise to Sensitivity in Private Data Analysis." In *Theory of Cryptography*, edited by Shai Halevi and Tal Rabin, 265–84. New York: Springer Berlin Heidelberg.
- Hu, Jingchen, and Claire McKay Bowen. 2023. "Advancing Microdata Privacy Protection: A Review of Synthetic Data Methods." *Wiley Interdisciplinary Reviews: Computational Statistics* 16 (1): e1636.
- JASON. 2022. "Consistency of Data Products and Formally Private Methods for the 2020 Census." US Census Bureau. <https://irp.fas.org/agency/dod/jason/census-privacy.pdf> (accessed on June 21, 2023).
- Kinney, Satkartar K., Jerome P. Reiter, and Javier Miranda. 2014. "Synlbd 2.0: Improving the Synthetic Longitudinal Business Database." *Statistical Journal of the IAOS* 30 (2): 129–35.
- Lohr, Steve. 2010. "Netflix Cancels Contest after Concerns Are Raised about Privacy." *New York Times*, March 13. <https://www.nytimes.com/2010/03/13/technology/13netflix.html>.
- Matthews, Gregory J., and Ofer Harel. 2011. "Data Confidentiality: A Review of Methods for Statistical Disclosure Limitation and Methods for Assessing Privacy." *Statistics Surveys* 5: 1–29.
- McClelland, Robert, Daniel Berger, Alyssa Harris, Chenxi Lu, and Kyle Ueyama. 2019. *The TCJA: What Might Have Been*. Washington, DC: Urban-Brookings Tax Policy Center.
- McKenna, Laura, and Matthew Haubach. 2019. "Legacy Techniques and Current Research in Disclosure Avoidance at the US Census Bureau." US Census Bureau Working Paper CED-WP-2019-005.
- Nagaraj, Abhishek, and Matteo Tranchero. 2023. "How Does Data Access Shape Science? Evidence from the Impact of US Census's Research Data Centers on Economics Research." NBER Working Paper 31372.
- Narayanan, Arvind, and Vitaly Shmatikov. 2008. "Robust De-anonymization of Large Sparse Datasets." In *2008 IEEE Symposium on Security and Privacy*, 111–25. Piscataway, NJ: IEEE.
- Rubin, Donald B. 1987. "Multiple Imputation for Nonresponse in Surveys." New York: Wiley.

- Snoke, Joshua, Claire McKay Bowen, Aaron R. Williams, and Andrés F. Barrientos.** 2024. "Incompatibilities between Current Practices in Statistical Data Analysis and Differential Privacy." *Journal of Privacy and Confidentiality*. <https://arxiv.org/pdf/2309.16703.pdf>.
- Taylor, Silke, Graham MacDonald, Kyle Ueyama, and Claire McKay Bowen.** 2021. *A Privacy-Preserving Validation Server Prototype*. Washington, DC: Urban Institute.
- Thompson, Stuart A., and Charlie Warzel.** 2019. "Twelve Million Phones, One Dataset, Zero Privacy." *New York Times*, December 19. <https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>.
- Williams, Aaron R., and Claire McKay Bowen.** 2023. "The Promise and Limitations of Formal Privacy." *Wiley Interdisciplinary Reviews: Computational Statistics* 15 (6): e1615.
- Williams, Aaron R., Joshua Snoke, Claire McKay Bowen, and Andrés F. Barrientos.** 2023. "Disclosing Economists' Privacy Perspectives: A Survey of American Economic Association Members on Differential Privacy and Data Fitness for Use Standards." Paper presented at the Data Privacy Protection and the Conduct of Applied Research: Methods Approaches and Their Consequences NBER conference, Cambridge, MA, May 4.

# When Privacy Protection Goes Wrong: How and Why the 2020 Census Confidentiality Program Failed

Steven Ruggles

**S**ince the first US Census in 1790, members of the public have expressed concerns about census intrusion into personal privacy. Public anxiety about census privacy appears to have peaked in the middle decades of the twentieth century and has generally declined since, except for a small uptick around the 2000 Census (Ruggles and Magnuson 2023). For almost two centuries, Census officials have responded to the public's privacy concerns with promises of confidentiality. This strategy has been ineffective; promises to keep personal information within the government do not address the core concern about government prying.

The Census Bureau maintains that strong confidentiality guarantees are essential for maximizing response rates to censuses and surveys, but little evidence supports this view. Indeed, experimental studies have consistently found that assurances of confidentiality actually *increase* concerns about confidentiality and *reduce* response rates to surveys (Berman, McCombs, and Boruch 1977; Frey 1986; Reamer 1979; Singer, Hippier, and Schwarz 1992). A Census Bureau analysis in the 1990s found that promises of confidentiality had no significant impact on response rates (Dillman et al. 1996).

The US Census Bureau recently implemented a new disclosure control strategy that marks a “sea change for the way that official statistics are produced and published” (Garfinkel, Abowd, and Powazek 2018, p. 136). The new disclosure control system adds deliberate error to every population statistic for every geographic unit smaller than a state, including metropolitan areas, cities, and counties.

Population data describing small geographic areas are essential for core political functions like drawing boundaries for state legislative districts and for the US House of Representatives. Towns, cities, counties, states, and the federal government use

■ *Steven Ruggles is Regents Professor of History and Population Studies, University of Minnesota, and Director of IPUMS, both in Minneapolis, Minnesota. His email address is [ruggles@umn.edu](mailto:ruggles@umn.edu).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.2.201>.

small area statistics for planning and policy purposes, ranging from decisions about delivery of public services to infrastructure needs. Moreover, economists and other social scientists rely on these demographic data to understand social changes and to evaluate policy outcomes. There is no historical precedent and no demonstrated need for introducing deliberate error into every population statistic for geographic units below the state level. These steps do nothing to allay public concern about the invasion of privacy by the government or about government misuse of data.

This article provides a critical history of the Census Bureau's evolving rationale for implementing this new approach to disclosure control. The Census Bureau justified the change by repeatedly claiming that the disclosure control system used for the 2010 Census revealed the confidential responses of millions of respondents. That claim, I contend, is entirely unsupported. Moreover, this new approach has undermined the utility of the 2020 Census data for scientific and policy research (Ruggles et al. 2019; Hotz et al. 2022; Muralidhar and Domingo-Ferrer 2023a). Ironically, the new system probably provides *less* confidentiality protection than the targeted disclosure control system it replaced (Kenny et al. 2023; Muralidhar et al. 2024).

## Census Bureau Disclosure Controls, 1970–2010

The Census Bureau has used several techniques to protect the confidentiality of respondents over the past half century (McKenna 2018). These traditional statistical disclosure control techniques introduced *uncertainty* into published data. The most important disclosure control method was “swapping,” which means exchanging a small percentage of households with similar households from a nearby area. In the 2000 and 2010 data, the swapped households had to match on their size and on the number of adults, but they could differ on any other characteristic. Every household had a chance of being swapped, but the algorithm especially targeted records with the most disclosure risk. Swapping focused on households containing persons with a unique combination of characteristics within their census “block.” Census blocks are the smallest geographic unit described by the census. There were about 6.4 million inhabited blocks in 2010, and the median person resided on a block with 109 people. Small census blocks pose the greatest disclosure risk, so the smaller the block, the higher the rate of swapping (Zayatz et al. 2009).

A concrete example of swapping, reported in the *New York Times* (Hansen 2018), involved the residents of Liberty Island, the site of the Statue of Liberty, which was a census block in 2010. The 2010 Census data show that the block included only one household with two people, a man in his 60s and a woman in her 40s, both identified as Asian. As it turns out, the *Times* had interviewed the actual couple in 2011 for an article about Liberty Island (Vadukul 2011). The real residents in 2010 were a 59-year-old husband and a 49-year-old wife, both identified as White. Without disclosure control, the published census would have revealed the confidential census responses of the only actual two residents on Liberty Island; in practice,

their responses were protected by swapping their household with a household from a nearby block that also consisted of two adults.

The process of data cleaning and editing introduces additional uncertainty into census data. When census information is missing or inconsistent—for a particular attribute of an individual, for an entire individual, or for a whole household—the Census Bureau substitutes (or imputes) information from a similar nearby person or household (Cantwell 2021). This imputation approach is conceptually similar to swapping, but the objective is to improve the data by reducing the number of missing cases, not to protect confidentiality.

Swapping introduces error into some counts, because the swapped households typically do not match on all characteristics. Prior to the 2020 Census, however, swapping and other disclosure control methods did not alter the counts of total population, voting age adults, housing units, or housing occupancy status at any geographic level. Swapping led to some error on other characteristics, but the Census Bureau concluded that “the impact in terms of introducing error into the estimates was much smaller than errors from sampling, non-response, editing, and imputation” (McKenna 2018, p. 24).

The traditional Census Bureau disclosure control strategy focused on ensuring that the *identity* of respondents—such as name, address, or Social Security number—cannot be inferred from Census publications. The Census Bureau implemented targeted disclosure controls to prevent reidentification attacks, so that an outside adversary cannot positively identify which person provided any particular response. The protections in place from 1970 through 2010 worked extremely well to meet that standard (Lauger, Wisniewski, and McKenna 2014). *Indeed, there is not a single documented case of anyone outside the Census Bureau revealing the responses of a particular identified person using data from the decennial census.*

## The Database Reconstruction Theorem

Despite the unblemished record of swapping-based disclosure control, in 2017 the Census Bureau decided that such methods were inadequate for the 2020 Census and that an entirely new approach was needed. The Bureau justified this decision by citing a “database reconstruction theorem” developed by computer scientists. The Census Bureau argued repeatedly: “The database reconstruction theorem is the death knell for traditional data publication systems from confidential sources” (for example, Abowd 2017, 2018a; Abowd et al. 2019).

The database reconstruction theorem, developed by Dinur and Nissim (2003), examined confidentiality protection in a hypothetical database. Dinur and Nissim envisioned a set of hospital records that include a secret binary medical condition (coded 0 or 1) along with a number of attributes, such as age and sex. In this exercise, researchers using the database are allowed to query the database and obtain the number of cases with a positive condition among patients with any given set of attributes; for example, they might ask for the number of positive cases among

women aged 30–39. To protect privacy, the hospital adds random noise to the result of each query. Dinur and Nissim proved that if the added noise is relatively small compared with the size of the database, it may be possible to reconstruct the original data given a sufficient number of random queries.

The hypothetical database query system envisioned by Dinur and Nissim (2003) bears little resemblance to published census tabulations (Muralidhar and Domingo-Ferrer 2023a, 2023b). The census does not allow random queries; instead, the Census Bureau decides in advance on a set of queries and presents the results in published tables. The overwhelming majority of possible queries cannot be answered by reference to these published tables. The Census Bureau’s swapping algorithm used prior to 2020 added some uncertainty to the published tables, but it was very different from the random noise envisioned by Dinur and Nissim. The database reconstruction theorem relies on random noise being injected *at the time of each query*; the underlying data can be revealed because the random noise varies with each query. By contrast, the uncertainty injected through swapping by the pre-2020 censuses was done only once. Moreover, the uncertainty added by the Census Bureau was not random; rather, it was targeted towards those cases likely to pose the greatest disclosure risk, such as residents of small blocks or persons with unique characteristics on their block.

Dinur and Nissim themselves point out that noise-infused data in a static database not subject to repeated queries *can* effectively protect confidentiality; they call this “*the CD Model*, where users get a ‘private’ version of the database (written on a CD)” (Dinur and Nissim 2003, p. 206). The CD model is essentially the same disclosure control strategy employed by the Census Bureau prior to 2020 (Muralidhar and Domingo-Ferrer 2023b). Thus, the database reconstruction theorem does not demonstrate a disclosure threat in published census tables.

## **The Census Bureau’s Database Reconstruction Experiment**

To demonstrate the threat of database reconstruction for census confidentiality, in 2016 the Census Bureau embarked on an experiment to reconstruct individual-level census responses using only the published tables from the 2010 census. The first public acknowledgment of the experiment came in a July 2018 presentation entitled “Staring Down the Database Reconstruction Theorem.” The presentation provided no actual results of the experiment, but it did claim that “the confidential micro-data from the hundred percent detail file can be reconstructed quite accurately” using only published tables (Abowd 2018a).

The first information on the results of the database reconstruction experiment appeared in a December 2018 article in the *New York Times*, the same article with the example of swapping on Liberty Island described above (Hansen 2018). Based on an interview with Census Bureau Chief Scientist John Abowd, the article explained that the goal of the reconstruction was to identify the age, sex, race, and Hispanic ethnicity for each individual in every census block in the country. Within each block, the Bureau generated a set of individual-level records consistent with

the published tables. The article reported: “By this summer, Mr. Abowd and his team had completed their reconstruction for nearly every part of the country. When they matched their reconstructed data to the actual, confidential records—again comparing just block, sex, age, race and ethnicity—they found about 50 percent of people matched exactly” (Hansen 2018, p. 7). Over the next eight months, Census Bureau staff made many additional public presentations and posted a 26-part “twee-torial” describing the threats exposed by the database reconstruction experiment (for example, Abowd 2018a, 2018b, 2018c, 2019; Abowd et al. 2019). These presentations provided no details about how the reconstructions worked, citing a need to keep the specifics confidential until a paper describing the reconstruction could be peer-reviewed for publication (Abowd et al. 2019).

The most detailed description of the Census database reconstruction experiment was the by-product of a lawsuit. On March 10, 2021, the State of Alabama filed a lawsuit against the Department of Commerce and the Census Bureau in US District Court, arguing that the proposed introduction of deliberate error to stymie database reconstruction—which would involve skewing every statistic below the state level—violated the right of the state to receive accurate tabulations of population (*Alabama v. Commerce*, 3:21-cv-211-RAH [2021]). (I served as an expert witness for the plaintiffs.) The case was dismissed on June 29, 2021, without prejudice to the underlying issues, on technical grounds related to questions of timing and standing. But the litigation resulted in a “declaration” from the US Census Bureau that included a twelve-page appendix providing a substantially more detailed description of the database reconstruction than had previously been available (Abowd 2021). My description below of the Census Bureau experiment draws mainly on this appendix.

The goal of the Census Bureau database reconstruction experiment was to convert tabular data describing the characteristics of census blocks in 2010 into microdata describing the characteristics of each individual on the block. Rearranging tabular data describing population characteristics into an individual-level format need not be complicated. Consider Table 1, a two-by-two table with information on eight people, broken down by race and sex. Table 2 shows exactly the same information as Table 1, but now expressed as individual-level microdata.

The individual-level characteristics that the Census Bureau attempted to reconstruct, based on published tables, were age, sex, race, and Hispanic ethnicity for the residents of each census block. The published 2010 Census includes a table (number P12A-I) that provides a breakdown of age by sex by race and Hispanic origin for every census block. As with Tables 1 and 2, it is simple to rearrange this table to obtain an individual-level dataset with block, age, sex, and race/ethnicity. Those microdata will be perfect replicas of the tabular data, although like the tabular data they include errors due to swapping, misreporting, nonresponse, and imputation. The detail of such constructed microdata, however, is limited. Most importantly, most ages are given in five-year bins (for example, 0–4, 5–9, and so on). Only five race groups are included, and other races are grouped into “some other race” or “two or more races.” Moreover, combinations of Hispanic origin and race are limited to Hispanic or Non-Hispanic and “White Alone.”

Table 1

**Hypothetical Population Data in Tabular Format**

	<i>White</i>	<i>Black</i>
Male	2	1
Female	3	2

Table 2

**Hypothetical Population Data in Microdata Format**

<i>Case number</i>	<i>Race</i>	<i>Sex</i>
1	White	Male
2	White	Male
3	White	Female
4	White	Female
5	White	Female
6	Black	Male
7	Black	Female
8	Black	Female

Thus, to reconstruct the database at the block level with greater detail, the Census Bureau turned to additional tables. One of these, Table P14, gives sex by single year of age for persons less than 20 years old on each block. If, for example, there is only one female aged 5–9 on a given block, then Table P14 can provide a specific age for that child. Another important table is PCT012A-N, which provides single years of age by race, ethnicity, and sex, but only at the census tract level. There are multiple blocks in every tract. If a tract contains just *one* person who is in a given age group, and who is of a specified race and sex, PCT012A-N can be used to assign a specific age to that individual.

Altogether, the Census Bureau used data from nine tables. The reconstruction experiment constructed a system of simultaneous equations consistent with the published tables. The investigators then solved the equations to create a set of hypothetical individual-level records consistent with all nine tables. The Census Bureau then compared those reconstructed microdata with the original census data by searching each block in the reconstructed data for cases that exactly match a case in the census on age, sex, race, and Hispanic ethnicity.

## Database Reconstruction: Results and Critique

The declaration from the Alabama court case provided precise match rates between the reconstructed data and the “true” census data. The Census analysts were able to find matches in the reconstructed data for 46.48 percent of the original unswapped Census microdata. How should this finding be interpreted?



The Census Bureau asserted that the experiment showed that “the micro-data from the confidential 2010 Hundred-percent Detail File (HDF) can be accurately reconstructed” using the published census tables (Abowd et al. 2019). Contrary to Bureau claims, an error rate of over 50 percent, however, is not an “accurate” reconstruction. Moreover, an outside attacker would have no means of knowing which of the reconstructed records matched someone in the real population, so the reconstructed data would not allow an outsider to positively identify any particular census respondent (Ruggles et al. 2019).

Perhaps unsurprisingly, given the task the Census Bureau had set for itself, most of the errors were because of age. As noted, the key table needed for reconstruction (P12A-I) provides only binned ages, mostly in five-year age groups, at the census block level. Individual years of age are only available at the census tract level. Inferring single years of age from five-year age bins was the biggest challenge facing the Census Bureau’s database reconstruction experiment.

As an illustration of the problem, consider Census Tract 5.01 in Laramie, Wyoming. In 2010, Tract 5.01 had 13 Black, non-Hispanic males in the 25–29 year age bin, spread over nine blocks. Muralidhar (2022) showed that just for these 13 individuals, there are 308,000 different ways to assign the exact ages from the tract-level data. The Census Bureau’s reconstruction procedure treats each of these 308,000 possibilities as equally likely and assigns the first one encountered. In the entire country, there are trillions of ways to “reconstruct” the population that are consistent with the published tabulations. The overwhelming majority of these reconstructions do not match the real population.<sup>1</sup>

How should we evaluate a match rate of 46 percent between the database reconstruction and the original census data? Any analysis of efficacy needs a control group. If a clinical trial found that 46 percent of patients who received a particular medication recovered from a disease, that would not prove that the medication is effective. One would also need to know what percent recovered among those receiving a placebo. Similarly, to evaluate the database reconstruction experiment, it is not sufficient to count the matches between the reconstructed population and the real population. Rather, we need a baseline to assess how much the reconstruction experiment outperforms a null model of random guessing.

To investigate the role of chance in the Census Bureau’s database reconstruction, Ruggles and Van Riper (2022) constructed a simple Monte Carlo simulation. The analysis estimated that a randomly selected person drawn from the 2010

<sup>1</sup>Recently the Census Bureau has improved its success metric for reconstruction by dropping their attempt to infer single years of age, and instead using the binned age groups (Hawes 2022; Abowd and Hawes 2023; Abowd et al. 2023). Using binned ages increased the match rate between the original census data and the reconstructed data from 46.5 to 91.9 percent. Because this reconstruction is effectively just a rearrangement of Table P012A-I into microdata format—much like the rearrangement shown above in Table 2—one might expect the match rate to be perfect. It is imperfect solely because of errors in race and ethnicity (these can arise for non-White Hispanics and for multiple-race respondents) and because of error introduced by swapping. Binned ages are generally not unique on their block and therefore pose much lower disclosure risk than do exact ages.

population would exactly match the age and sex of somebody on a randomly selected block 52.6 percent of the time. One would therefore expect the Census Bureau to be “correct” on age, sex, and block most of the time even if they had never looked at the tabular data from 2010 and had instead just assigned ages and sexes to a hypothetical population at random.

The simulation by Ruggles and Van Riper (2022) does not factor in race or ethnicity, but because of high residential segregation, most blocks are highly homogeneous with respect to race and ethnicity. If we assign everyone on each block the most frequent race and ethnicity of the block, then race and ethnicity assignment is correct in 77.8 percent of cases (Manson et al. 2022). Using that method to adjust the random age-sex combinations described above, 40.9 percent of cases would be expected to match on all five characteristics to a respondent on the same block. That does not differ greatly from the Census Bureau’s reported 46.5 percent match rate for its reconstructed data (Abowd 2021, App B, p. 3). The analysis suggested that among the minority of cases where the Census Bureau did find a match between their hypothetical population and a real person, most of those matches would be expected to occur purely by chance.

Recently, a group of Census Bureau analysts argued that the Ruggles and Van Riper analysis “was severely flawed” (Jarmin et al. 2023). To estimate a baseline match rate, Ruggles and Van Riper (2022) estimated the likelihood that a randomly-selected individual would be present on a randomly-selected block. Jarmin et al. argue that Ruggles and Van Riper should instead have closely modeled the matching procedure used by the Census Bureau, and they provided previously undocumented details of that algorithm. It is straightforward to adjust the Monte Carlo simulation to conform to the Census Bureau’s specifications (Ruggles and Van Riper 2023). The modification reduces the match rate obtained by random assignment, but the overall conclusion is unchanged. The revised simulation still shows that the great majority (68.6 percent) of the matches between the Census Bureau’s reconstructed data and the original census data would be expected with random assignment of age and sex and systematic assignment of race and ethnicity. Thus, relatively little additional information about census respondents is gained by drawing on the published census tables in the reconstruction exercise.

The main point is that a match rate is not a valid indicator of disclosure risk. Dick et al. (2023)—prominent advocates of the database reconstruction theorem—endorse this core argument. They write that Ruggles and Van Riper (2022) “does raise the important question of choosing appropriate baselines and what we should measure, beyond top-level reconstruction rates, to indicate when we should view reconstruction attacks as worrisome” (Dick et al. 2023, p. 2).

## **The Census Bureau Reidentification Experiment**

Reconstruction and reidentification are very different, and only reidentification poses a meaningful risk to census respondents (Muralidhar and Domingo-Ferrer

2023a). The reconstructed data produced by the Census Bureau contain no personal identifiers. A disclosure attack requires reidentification, meaning that the attacker must accurately link the reconstructed data to an external database that reveals people's names.

The standard procedure in reidentification studies is to look for unique combinations of variables in individual-level research data (such as survey data) that match the same variables in an external source. Those unique matches are considered vulnerable to reidentification. A reidentification is confirmed only if the identity of the person—ordinarily verified by name—is the same in the research data and in the external data (McKenna 2019; McKenna and Haubach 2019). Thus, for example, a medical record in an anonymized research dataset containing exact date of birth, sex, and zip code might be uniquely matched to public voter records that include the same three variables, thus revealing the name and address of the patient (Barth-Jones 2012). That reidentification would be “confirmed” only if the name on the internal medical records matched the name in the voter file.

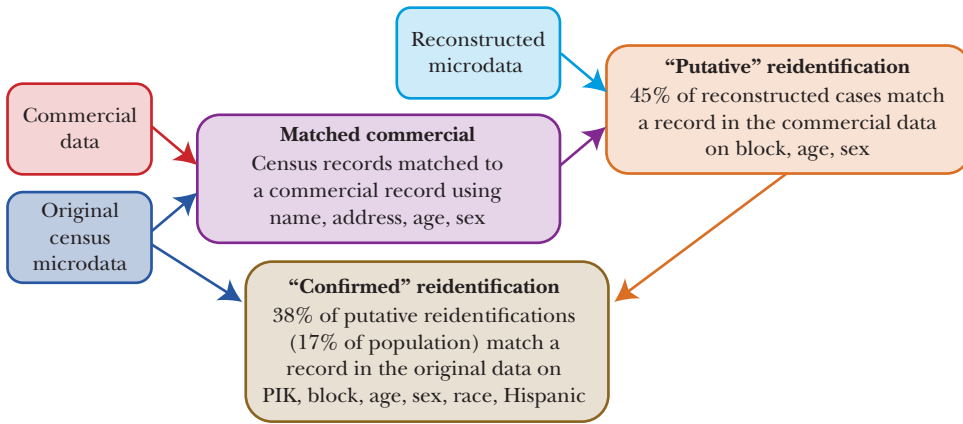
Initially, Census Bureau presentations on database reconstruction consistently maintained that although the reconstruction was accurate, the risk of reidentification was small (for example, Abowd 2018a, 2018b; Abowd et al. 2019). In February 2019, the Census Bureau abruptly reversed course, asserting that they had established “confirmed reidentifications” for 17 percent of the 2010 population, exposing the confidential responses of 53 million identified respondents (Abowd 2019).

For the reidentification component of the Census experiment, the Census Bureau used financial and marketing data from five commercial vendors: Experian, Targus, Veteran Service Group of Illinois, InfoUSA, and Melissa. The information in the commercial files was drawn from credit agency reports, magazine subscription records, utility bills, property tax records, voter registration data, and other sources. These files consistently provided information on name, address, age, and sex, and sometimes other variables such as race and Hispanic origin. Shortly after the 2010 Census, the Census Bureau merged all the files, removed the duplicates, and matched them to the 2010 Census returns using the information on name, address, age, and sex. At that time, the goal of the study was to assess the feasibility of using commercial data to improve or substitute for enumerated census data (Rastogi and O'Hara 2012).

Figure 1 diagrams the reidentification process. Each case in matched commercial data should have the same name, address, age, and sex as the corresponding record in the original census file. The Census Bureau assigned a unique numeric code called a “Protected Identification Key” (PIK) to each individual; for each matched individual, an identical PIK appeared in both the original census file and in the matched commercial file.

The first step in the reidentification experiment was to search the matched commercial file for a record that matched each record in the reconstructed microdata on block, sex, and age. This was done by “looping through all the records in the reconstructed microdata file produced from the reconstruction, find the first record in the source file [the commercial data] that matches exactly

Figure 1

**The Census Bureau Reidentification Experiment**

Source: Author's creation.

on block, sex, and age” (Abowd 2021: App. B, p. 7). This match yields the “putative” reidentifications, which represent 45 percent of the population.<sup>2</sup> The Census Bureau then attached the PIK code from the commercial data to the putative reidentifications.

As the Census declaration in the *Alabama v. Commerce* lawsuit noted: “Putative reidentifications are not necessarily correct” (Abowd 2021: App. B, p. 8). The putative reidentifications are just cases where a row of the reconstructed microdata has the same block, age, and sex as an individual in the commercial database. If there are multiple people in the reconstructed microdata of the same age and sex—as often occurred—the first one encountered is considered the putative reidentification.

The Census Bureau considers putative reidentifications “confirmed” if they match someone in the original census on the PIK number, race, and Hispanic origin in addition to block, age, and sex. A core problem, however, is that the reconstructed data do not have a PIK; the PIK in the putative file is copied from the commercial data. The Census Bureau assigned PIKs to both the commercial data and the original census data based on name, address, age, and sex (Wagner and Lane 2014). Because the PIK code in the putative data comes from the

<sup>2</sup>This is potentially confusing because the 45 percent putative match rate is so close to the 46 percent match rate between the reconstructed data and the original census on block, age, sex, race and Hispanic origin. This is purely a coincidence; the two matched sets are overlapping but distinct.

commercial data, it ordinarily should match a PIK in the original census data. Those cases should also match the census on block, age, and sex, because address, age, and sex were used to assign the PIKs in the first place. Using PIK, block, age, and sex to confirm putative matches is therefore circular: all a match really means is that someone in the commercial data also appears in the census.

To be designated as confirmed, putative cases must also match the original census data on race and Hispanic origin (in addition to block, age, sex, and PIK). As noted earlier, because of the high residential segregation in the United States, most blocks have little diversity with respect to race and ethnicity. Indeed, 78 percent of the population from the 2010 Census identified with the most frequent race and ethnicity found on their block. One would therefore expect the great majority of putative reidentifications to be confirmed just because most people's race and ethnicity matches the race and ethnicity of most other people on their block.

Given the circular reasoning, one would anticipate a high rate of confirmation of putative reidentifications. The Census Bureau, however, reports strikingly low confirmation rates. The confirmed reidentifications comprise just 38 percent of the putative reidentifications, or 17 percent of the whole population (that is, 38 percent of the 45 percent of putative reidentifications were confirmed). The most plausible explanation for such a low confirmation rate is that the commercial data are highly incomplete and inaccurate.<sup>3</sup>

The reidentification procedure used by the Census Bureau may sound superficially similar to a standard reidentification study, but it is actually very different. Reidentification studies use names to confirm identity. In the Census Bureau experiment, the external source—the matched commercial file—had already been matched to the original census data using name, address, age, and sex prior to the experiment. The Census Bureau could not use independently acquired information on name, as would be done in a standard reidentification study, because the reconstructed data do not include independent information on name. By the standards and procedures used in prior Census Bureau reidentification studies, none of the reidentifications of the 2010 reconstructed data would be considered confirmed (McKenna 2019; McKenna and Haubach 2019).

<sup>3</sup>In the Census Bureau's newest reidentification estimates, they use binned ages as described in note 1. Then, instead of using the relatively low-quality commercial data, they use the original census data to define putative reidentifications, on the grounds that an external attacker might somehow have access to extremely high quality data (Abowd and Hawes 2023; Abowd et al. 2023). The confirmed match rate (in which there is a match between the putative reidentifications and the original data on Protected Identification Key, block, sex, age group, race, and ethnicity) is 75.5 percent. Given that this exercise is effectively matching the original individual-level census data back to itself, that confirmation rate seems surprisingly low; by definition, 100 percent of the putative rows will match the original census data on PIK, block, sex, and age group. The mismatches can result only from incorrect race or ethnicity or from swapping.

## Implications of the Reidentification Experiment

Does the Census Bureau's reidentification experiment demonstrate a realistic threat to confidentiality? The Census Bureau argues that their reidentification could reveal a respondent's confidential responses to the race and ethnicity questions (Abowd 2021; Garfinkel 2023). The idea is that an external attacker could infer an individual's race and Hispanic origin by matching reconstructed data on age, sex, and block to an external source that revealed identity, thus making a putative reidentification.

Based on the Census Bureau's own analysis, this approach would be highly inaccurate; 62 percent of the putative reidentifications were definitely incorrect. The exercise would also be pointless: as Francis (2022) pointed out, the Census Bureau's elaborate attack strategy is far less reliable than simply inferring race and ethnicity based on the characteristics of the block. By assigning the modal race and ethnicity of the block, one can accurately guess race and ethnicity in about 75 percent of cases. For the 11 percent of the population residing on perfectly homogeneous blocks, one can infer race and ethnicity with perfect accuracy (except for uncertainty introduced by swapping and imputation).<sup>4</sup>

In a recent working paper, Census Bureau analysts acknowledge that Francis (2022) was correct in the great majority of cases: for the bulk of the population, the database reconstruction and reidentification exercise could not help an outsider guess race and ethnicity any better than guessing the modal race and ethnicity of the block (Abowd et al. 2023). The Census Bureau analysts argue, however, that reconstruction and reidentification are effective for a particular subset of the reconstructed data: reconstructed rows that do not have the modal race and ethnicity of their block and that are unique on their block with respect to binned age and sex. They call these cases "nonmodal uniques." Because these cases are unique on their block with respect to binned age and sex, their race and ethnicity can usually be read directly from Table P12A-I; no reconstruction is needed. Among this subset, the paper reports that about one in six cases (representing 0.19 percent of the total population) matched someone in the commercial data on binned age and sex. These are the "putative reidentifications." The authors conclude that the putative reidentifications of nonmodal uniques "definitively show" that the published tables "result in confidentiality breaches" (Abowd et al. 2023, p. 47).

Without access to confidential internal census data, an outside attacker on the 2010 Census would have no means to gauge the reliability of attempted reidentifications of the nonmodal uniques. Because swapping targets cases with unique characteristics, a potential attacker would likely assume an exceptionally high error rate for this group. Abowd et al. (2023) have now revealed a somewhat

<sup>4</sup>Jarmin et al. (2023) complain that Francis (2022) ignores the effects of swapping and imputation in 2010 (even though the Census Bureau's reconstruction also ignores swapping and imputation). The results in *Alabama v. Commerce* (3:21-cv-211-RAH [2021]) suggest that the impact of swapping is small compared with the extremely high error rate in the database reconstruction.

lower error rate for the putative nonmodal uniques than one might expect.<sup>5</sup> Nevertheless, many of the race and ethnic inferences are still incorrect, and an outside attacker has no means of confirming whether any particular inference is true.

The database reconstruction and reidentification as implemented by the Census Bureau posed no realistic threat to the confidentiality of the 2010 Census. As explained above, the database reconstruction theorem does not apply to published census tables. The Census Bureau's reconstructed data were decidedly inaccurate, performing little better than a random number generator. Even if the Census Bureau could somehow improve the quality of the reconstruction, the only thing to reconstruct would be the swapped version of the data, so there would always be uncertainty about whether any given row of the reconstruction appeared on the block in real life. Reidentification studies ordinarily require a match on name or another reliable identifier to confirm any putative reidentification (McKenna 2019). The reconstructed data have no such identifiers; therefore, the only confirmed links possible are between the commercial data and the census, not between the reconstructed data and the census.

In a 2019 blog post, the Acting Census Bureau Director acknowledged: "The accuracy of the data our researchers obtained from this study is limited, and confirmation of re-identified responses requires access to confidential internal Census Bureau information . . . more than half of these matches are incorrect, and an external attacker has no means of confirming them" (Jarmin 2019). Five years later, that assessment has proven valid: the kind of reconstruction and reidentification attack used in the Census Bureau's experiment does not allow positive identification of any census respondents.

## **Differential Privacy and Noise-Infused Tabular Census Data**

In September 2017, the Census Bureau announced to the Census Scientific Advisory Committee that the 2020 Census would abandon traditional disclosure controls such as swapping and instead use a new "differentially private" disclosure avoidance system to ensure confidentiality, in response to the threat allegedly posed by the database reconstruction and reidentification (Garfinkel 2017).<sup>6</sup> The application of

<sup>5</sup>In the past, the Census Bureau kept such statistics secret, because they might enable an attacker to gauge the reliability of an attempted reidentification. Abowd et al. (2023) for the first time reveal the error rate on race and ethnicity for the subset of nonmodal uniques who match someone in the commercial data on age and sex.

<sup>6</sup>In describing the plan, Dajani et al. (2017) describes an "agreement with the Department of Justice" under which the "Census Bureau will provide exact counts at the Census block level" for the total population and the adult (age 18+) population. In August 2020, the Census Bureau posted a revised version of that paper that "supersedes the 2017 version" (Abowd et al. 2020). The paper specifies that exact population counts would not be provided at the block level, nor at the tract, county, or city levels. The only exact population counts in the 2020 Census would be for entire states; every statistic below the state level would have random errors added. In the new iteration, the agreement between the Census Bureau and the Justice Department was not mentioned.

differential privacy to census data represents a radical departure from established Census Bureau precedents.

Instead of guaranteeing that census responses cannot be tied to particular individuals, differential privacy guarantees that the presence or absence of any individual case from a database should not significantly affect any database query. The requirement that database outputs do not significantly change when any individual's data is added or removed has profound implications. In effect, under differential privacy, it is prohibited to reveal *characteristics* of an individual even if the *identity* of that individual is effectively concealed. In other words, differential privacy is more concerned about database reconstruction than about reidentification. This redefinition of privacy makes disclosure control significantly more challenging compared with the traditional Census Bureau focus on preventing positive identification of respondents.<sup>7</sup>

The big advantage of the new definition of privacy is that it is relatively simple to formalize, and that formalization yields a metric summarizing a database's level of "privacy" in a single number. The core metric used in the differential privacy literature is epsilon ( $\epsilon$ ), which is often referred to as the "privacy-loss budget." When  $\epsilon$  is large, noise infusion is limited and confidentiality protection (under the new definition) is low; when  $\epsilon$  is small and near-zero, noise infusion is large, and disclosure control is high. Of course, adding a high degree of randomness to many data parameters can also make census data less useful, or not useful at all, for purposes of public policy and research (Dwork et al. 2006; Bambauer, Muralidhar, and Sarathy 2014).

Based on the theory of differential privacy, the Census Bureau implemented an elaborate procedure to inject random noise into tabular census data. The noise-infusion algorithm is clearly documented in reports of the National Academies of Science, Engineering, and Medicine (Committee on National Statistics 2020, 2023a; Sullivan and Cork 2022). There are five main steps. The process begins by constructing tabular data from the individual-level census returns. Second, a controlled amount of random statistical noise is added to each cell of the tables; the amount of noise used varies according to geographic level and variable. The third step is post-processing, to make sure that the tables do not include logical impossibilities (such as negative population counts) and are internally consistent. Fourth, the Census Bureau converts the processed tables into microdata using the same database reconstruction method they had used to support a need for differentially

<sup>7</sup>Abowd (2018c: 15) argues that there is a legal basis for the new definition of privacy: "Re-identification risk is only one part of the Census Bureau's statutory obligation to protect confidentiality. The statute also requires protection against exact attribute disclosure." This suggests that confidentiality law prohibits revealing respondent characteristics even if the respondent's identity is protected. The statute in question prohibits "any publication whereby the data furnished by any particular establishment or individual under this title can be identified" (Title 13 USC. § 9, Public Law 87-813). Previously, the Census Bureau interpreted "particular individual" to be an individual whose identity is known.



private census data. Finally, the reconstructed microdata are tabulated to prepare standard census tables for public release.<sup>8</sup>

To test this data production algorithm, the Census Bureau released, between October 2019 and April 2021, a series of six “demonstration files” based on the 2010 Census (Van Riper, Kugler, and Schroeder 2020–2023). These demonstration files allowed external investigators to assess the usability of the noise-infused data for research and public planning purposes. Most external analysts concluded that all these demonstration files were unfit for critical research and policy applications. Because of the post-processing step, the disclosure avoidance system not only introduced unacceptable levels of random error for many applications of the census but also introduced systematic biases. For example, the demonstration files systematically reduced the size of urban and suburban populations and increased rural populations. The demonstration files also reduced minority populations where minorities are highly concentrated and increased them in areas with lower minority concentration (Santos-Lozada, Howard, and Verdery 2020). The test data undercounted mixed-race and mixed-partisan precincts, posing serious concerns for redistricting (Kenny et al. 2021). They substantially distorted net migration estimates, making migration rate calculations unusable in about half of counties (Winkler et al. 2022). They distorted COVID-19 mortality rates and measures, sometimes causing mortality rates to exceed 100 percent (Hauer and Santos-Lozada 2021). The demonstration data also introduced systematic error into measures of residential segregation (Asquith et al. 2022).

Newspaper reports and social media piled on, highlighting anomalies and internal inconsistencies in the demonstration files (for example, Capps 2021; Menger 2021; Schneider 2021; Wines 2022). In some cases, the noise infusion algorithm made occupied neighborhoods vanish; in other cases, the algorithm populated uninhabited blocks. The data included hundreds of thousands of “Lord of the Flies Blocks” consisting entirely of children with no adults present; “Mermaid Blocks” consisting of people residing in vacant housing units located in lakes and rivers; and “Ghost Blocks” with occupied homes but zero population.

When the Census Bureau announced in June 2021 the final specifications for the electoral redistricting data, they shocked the data user community by specifying far less noise than they had used for the demonstration files. As a result, the corresponding privacy-loss budget  $\epsilon$ —the statistic summarizing the amount of error introduced—was also many times higher than is ordinarily contemplated by privacy researchers. Recall that when the summary metric  $\epsilon$  is large, noise infusion is limited, and confidentiality protection is low. The range of  $\epsilon$  in the differential privacy literature generally runs from 0.01 to 5.0, but many analysts argue that, to guarantee privacy,  $\epsilon$  should not greatly exceed 1.0 (Lee and Clifton 2011; Dwork

<sup>8</sup>The procedure was used for the initial redistricting file (PL. 94-171) and for the “demographic and housing characteristics” file (DHC). The detailed demographic and housing characteristics file (DDHC-A) and the yet-to-be released DDHC-B file and supplemental DHC file use an entirely different noise-infusion algorithm developed by an outside contractor (US Census Bureau 2023a).

2011). Several years ago, Apple announced that it would use a differential privacy approach to protecting personal data, but with a value of  $\varepsilon = 14$ . Frank McSherry, one of the co-inventors of differential privacy, remarked at the time that “anything much bigger than one is not a very reassuring guarantee.” He argued, “Apple has put some kind of handcuffs on in how they interact with your data. It just turns out those handcuffs are made out of tissue paper.” McSherry went on to describe Apple’s disclosure controls as “relatively pointless” (as reported in Greenberg 2017; Domingo-Ferrer, Sánchez, and Blanco-Justicia 2021).

In response to the scholarly and public criticisms of the demonstration files, the Census Bureau used an even higher  $\varepsilon = 19.61$  for the final specifications of the redistricting data file. Because the scale of the privacy loss budget is exponential, this privacy-loss budget for the redistricting file was “exponentially higher” than the highest budget used in any of the demonstration files that had been released earlier (US Census Bureau 2021). Accordingly, the Census Bureau’s implementation of differential privacy provides minimal data security. Indeed, this new approach probably provides *less* confidentiality protection than the traditional disclosure controls (mainly swapping) used by the Census Bureau before 2020.

Since the release of the 2020 data, the Census Bureau has produced new demonstration files for the 2010 Census designed using the differentially private disclosure controls actually used for the 2020 Census. The demonstration files allowed external analysts to assess the impact of the new disclosure controls on the usability of the 2020 data. The studies conducted to date suggest that the reduction in the level of noise has improved data quality, but also that the noise-infused data remain substantially inferior to the data based on disclosure avoidance techniques originally used in 2010, and the noise-infused data remain unusable for some applications. For example, one analysis found that the disclosure control introduced errors exceeding 5 percent in the number of young children for 27 percent of school districts (Committee on National Statistics 2023b, p. 34). Another analysis revealed dramatic differences in household composition among the elderly in the differentially private data (Committee on National Statistics 2023b, p. 44). Mueller and Santos-Lozada (2022) demonstrated that even with  $\varepsilon = 19.61$ , the noise infusion introduced an unacceptable level of error for small populations, especially non-Whites and inhabitants of rural areas, raising serious questions about the validity of the approach. The final redistricting data still include many internal inconsistencies, including 91,000 blocks with occupied housing units but no people, 101,000 blocks with occupied households but only children present, and 309,000 blocks with people in vacant housing units.

The Census Bureau committed to using differentially private disclosure control in 2017, before the algorithms for adding noise were developed and tested. The design of the system and software was still at an early stage, and implementation ultimately proved far more difficult than anticipated. Indeed, at the time of this writing in early 2024, the Census Bureau is still developing algorithms needed to produce critical tables for the 2020 Census. The implementation of differentially private disclosure controls has led to a years-long delay in the delivery of census tabulations.

When it became clear in June 2021 that the initial algorithm was producing data unfit for use, at the last minute the Census Bureau reduced the noise infusion to levels far below the normal standard required for disclosure protection. Despite the reduced noise, many academic users remain dissatisfied with data quality, and planners and policymakers have lost faith in the reliability of the 2020 statistics for small areas. As more tables were released, the Census Bureau was forced to reduce the amount of noise still further and to raise the privacy-loss budget to the previously unheard-of level of  $\epsilon = 39.9$  (US Census Bureau 2022). To maintain some degree of disclosure control, the Bureau has also substantially reduced the number of published statistics from the 2020 Census relative to the 2010 Census.

There is no evidence that the Census Bureau's current implementation of differential privacy, using extremely small levels of random noise, protects confidentiality as well as traditional statistical disclosure control. An evaluation commissioned by the Census Bureau concluded that the 2020 disclosure control as implemented "does not provide any comforting guarantees" of confidentiality (JASON 2022, p. 116) and "not enough is known about whether the privacy mechanisms as implemented are sufficient to mitigate the disclosure risks that motivated adoption of formal privacy" (p. 9). Kenny et al. (2023) found that noise-infused data substantially increase the odds of correctly guessing the racial identification of particular respondents. Accordingly, the disclosure control system implemented by the Census Bureau not only reduces the utility of the data, but also provides minimal confidentiality protection (Muralidhar et al. 2024).

## **Synthetic Microdata**

The Census Bureau is on the brink of another disclosure control blunder. The American Community Survey (ACS) Public Use Microdata Sample (PUMS) provides annual samples describing the demographic and economic characteristics of about 3.4 million individuals and 1.3 million households. The ACS PUMS continues a series of large microdata samples produced by the Census Bureau since 1962, providing detailed characteristics of individual respondents. To protect confidentiality, the Census Bureau does not identify places with less than 100,000 population, and the Bureau uses the traditional tools of swapping, top-coding of high values for continuous variables such as income and age, and perturbation of some ages. The ACS PUMS is also protected because it is a sample of just 1 percent of the population. Even if one finds a unique match of certain characteristics between a respondent in the survey and an individual in an external source, one can never be sure that the match is unique in the population as a whole.

The risk of reidentification is very low; the most recent reidentification study of the American Community Survey microdata showed that 0.017 percent of respondents were vulnerable to possible reidentification, but 78 percent of those putative reidentifications were false, and an outsider would have no means of determining which ones were correct (Ramachandran et al. 2012). There is no documented case

in which a response to the American Community Survey has ever been linked to an identified person by someone without access to internal Census Bureau data.

Despite this strong record of confidentiality protection, the Census Bureau is planning to replace the American Community Survey microdata with “fully synthetic” data to bolster confidentiality (Rodríguez 2021; Daily 2022). The idea of fully synthetic microdata is to develop models describing the interrelationships of all the variables in the data and then use random draws to construct a simulated population consistent with those models (Abowd et al. 2020).

Synthetic data captures relationships between variables only if those relationships have been anticipated in advance and intentionally included in the models. Accordingly, synthetic microdata are poorly suited to studying unanticipated relationships, which impedes new discovery. The large size of the American Community Survey means that it is possible to study small population subgroups, but synthetic data cannot capture all the ways in which interrelationships among variables can vary across subgroups. For example, fully synthetic ACS data would certainly incorporate a general relationship between education and income, but would not assess that relationship separately for every population subgroup. The relationship between education and income might be very different for American Indians in South Dakota compared with Asian Indians in the Queens borough of New York City. Nor would synthetic ACS data capture the myriad possible interrelationships among the characteristics of different family members, such as the relationship between a person’s education and their spouse’s education.

These limitations are important, because the American Community Survey microdata are among the most intensively used scientific data sources in the world and are bedrock resources for demographic and economic research. Tens of thousands of academic researchers, planners, and policy makers rely on the ACS, and according to Google Scholar they generate over 10,000 publications per year. Common topics of analysis include poverty, inequality, immigration, internal migration, ethnicity, residential segregation, disability, transportation, fertility, marriage, occupational structure, education, and family composition. If public use data become unusable or inaccessible, the quantity and quality of research about US policies, the economy, and social structure will decline dramatically.

The Census Bureau acknowledges that the synthetic data will not be reliable enough to support research applications. Consequently, a central element of the plan is to provide “validation” services to researchers. In theory, the researchers will conduct their analyses on the synthetic data and then submit their code to the Census Bureau, which will run the code against the internal “true” data. The Census Bureau will then put the output through disclosure control procedures and give the results back to the researchers (Abowd et al. 2020).

This strategy has multiple flaws. Investigators need access to real data for exploratory analyses to discover relevant variables to incorporate in their analyses, not just

for validation. Moreover, as Muralidhar (2023)<sup>9</sup> points out, the validation procedure would potentially disclose more information than the current system does, because the validation server would use the original unswapped data; given a sufficient number of validation queries, the underlying data could be subject to exact reconstruction as in the Dinur and Nissim (2003) scenario. The plan is impractical; the Census Bureau lacks the resources to provide this validation service at scale. Executing the plan would be expensive, partly because all the output from the validations must undergo full disclosure review before being released to the researcher. To provide service comparable to the current usage of the American Community Survey microdata, the Census Bureau would have to validate hundreds of thousands of analyses per year.<sup>10</sup> If the ACS microdata are replaced with fully synthetic data, that would represent the most damaging loss in access to data describing the US population in the history of our statistical system. The Census Bureau has not provided evidence of disclosure risk from existing practices that would justify this radical change.

## Balancing Privacy and Usefulness of Data

Differentially private noise injection was an inappropriate disclosure control choice for the census. The Census Bureau never attempted to weigh realistic measures of disclosure risk under alternative disclosure control methods against the harm of producing an unreliable census (Hotz et al. 2022). The noise-injection algorithm is a blunt instrument that adds deliberate error to every statistic below the state level. Differentially private noise injection is indiscriminate; unlike swapping, it does not target the most vulnerable respondents. Even the total population of New York City is perturbed. This is pointless: tabular data for large populations do not need disclosure control, because there are no cases with unique combinations of characteristics.

For small areas, disclosure control for statistical publications is essential. Without disclosure control, the census responses of the only two people residing on Liberty Island in 2010 would have been compromised, as would the responses of millions of others. Anyone with a unique set of characteristics on a census block is potentially identifiable. Swapping is an attractive tool for disclosure control because it does minimal damage to accuracy, preserves the counts of the number of people and the number of adults at every level of geography, and can be effectively targeted to focus on people at high risk of disclosure. It is not necessary to swap every case with a unique set of characteristics; one need only swap a sufficient proportion to

<sup>9</sup> This was pointed out by Muralidhar in a personal communication in 2023.

<sup>10</sup> The Census Bureau has suggested that users who need reliable American Community Survey microdata could gain access through the network of Federal Statistical Research Data Centers (FSRDCs). This would be challenging: there are currently only a little over 200 FSRDC projects using census data across the entire network, and disclosure review is already a significant bottleneck. Providing broad access through the data centers would require expansion of the capacity of these centers by multiple orders of magnitude.

create uncertainty, so that an outside attacker can never be confident of a particular respondent's identity.

Census law prohibits the positive identification of respondents. The law does not prohibit publication of statistics that can help an attacker guess someone's characteristics based on place of residence. Indeed, all census statistics improve the chances of guessing respondent characteristics. If I know only that someone lived in Vermont on census day in 2010, I can guess with 94.4 percent confidence that the person identified as White and non-Hispanic (Manson et al. 2022). That is far more precise than estimates of race and ethnicity responses based on the Census Bureau reconstruction and reidentification experiment. In fact, even the tiny subgroup with the highest reidentification accuracy—described by the Census Bureau as the “nonmodal uniques”—has about the same error rate (Abowd et al. 2023).

An external report on the new disclosure controls commissioned by the Census Bureau argues that “the risk that matters is if the released data allows an adversary to make inferences about an individual's characteristics with more accuracy and confidence than could be done without the data released by the Census Bureau” (JASON 2022, p. 114). This interpretation—if taken literally—would have profound consequences. If the Census Bureau could not publish anything that improves the chances of guessing someone's characteristics, then all publication of population characteristics—even at the national level—would be prohibited.

To reduce the risk of disclosure of census responses, the Census Bureau could take steps to reduce the number of tabulations with unique combinations of characteristics. Many census blocks are extremely small: in 2010 there were 195,339 blocks with a single person, and 76 million people resided in blocks with fewer than 50 people (Manson et al. 2022). There are not many use cases for such tiny blocks; they could be consolidated into neighboring blocks to reduce disclosure threats. If the Census Bureau consolidated the blocks with fewer than 50 people by merging them with other blocks, that would eliminate 68 percent of the cases that have unique combinations of age, sex, and block. If blocks smaller than 100 were eliminated, that would take care of over 90 percent of the unique age-sex-block combinations.

A similar strategy for minimizing disclosure risk would be reducing the detail of Census tables provided at the block level. Although the Census Bureau has greatly reduced the number of tables to be produced for the 2020 Census, the Bureau oddly did not eliminate the key table that underlies their database reconstruction, the P12A-I table that shows age by sex by race and Hispanic origin for every census block. Surprisingly, the Census Bureau is substantially expanding the detail provided in this table by tabulating a full set of the interactions between Hispanic origin and race, as well as many race combinations (US Census Bureau 2023b). The additional information makes database reconstruction easier and more accurate. The Census Bureau has not explained the reasons for providing this extra detail.

When uncertainty is added to protect confidentiality—either through swapping or through targeted noise injection—that should not affect the total counts of population. Population size is by far the most important census statistic for small

areas and should be reported accurately; revealing true population counts need not compromise anyone's privacy. Accordingly, swapping or noise injection should be designed to add uncertainty to census responses on detailed age, sex, race, ethnicity, or family relationship, without altering the total population counts.

There is a compelling case for the social benefit of *broadening* access to reliable data from federal agencies (Commission on Evidence-Based Policymaking 2018). Access to high-quality data makes social research and policy formation more reliable, less expensive, and more reproducible. The Foundations for Evidence-Based Policymaking Act of 2018 (2019) requires each federal agency to develop open data plans and make federal data publicly available by default (Commission on Evidence-Based Policymaking 2018). In the ensuing years, however, federal agencies have *reduced* public access to reliable data because of unproven worries about confidentiality. To ensure that government agencies do not curtail broad access to rich and reliable data without clearly demonstrated need, we need further legislation in two areas. First, we need to clarify that disclosure control laws protect against positive identification of particular respondents, and do not prohibit publishing population characteristics. Second, we should mandate that agencies do not withdraw access to data without balancing the social cost of losing data against realistic measures of the risk of harm to individuals (Hotz et al. 2022).

■ *Work was supported by Alfred P. Sloan Foundation grant G-2019-12589, "Implications of Differential Privacy on Decennial Census Data Accuracy and Utility," and Eunice Kennedy Shriver National Institute of Child Health and Human Development grant P2C HD041023, "Minnesota Population Center." I am grateful for the extensive feedback I received from Margo Anderson, Andrew Beveridge, Josep Domingo-Ferrer, Paul Francis, V. Joseph Hotz, Miriam L. King, Diana Magnuson, Robert Moffitt, Krish Muralidhar, Theresa Sullivan, and David Van Riper, as well as JEP editors Heidi Williams, Nina Pavcnik and Erik Hurst. I especially appreciate the exceptional efforts of Managing Editor Timothy Taylor, who reorganized and largely rewrote the entire article.*

## References

- Abowd, John M.** 2017. "Research Data Centers, Reproducible Science, and Confidentiality Protection: The Role of the 21st Century Statistical Agency." Presentation, Summer DemSem, Wisconsin Federal Statistical RDC, June 5, 2017. <https://www2.census.gov/cac/sac/meetings/2017-09/role-statistical-agency.pdf>.
- Abowd, John M.** 2018a. "Staring-Down the Database Reconstruction Theorem." Presentation, Joint Statistical Meetings, Vancouver, BC, July 30, 2018. <https://www.census.gov/content/dam/Census/newsroom/press-kits/2018/jsm/jsm-presentation-database-reconstruction.pdf>.
- Abowd, John M.** 2018b. "The U.S. Census Bureau Adopts Differential Privacy." Speech, 24th ACM

- SIGKDD Conference on Knowledge Discovery and Data Mining, London, August 23, 2018. <https://dl.acm.org/doi/10.1145/3219819.3226070>.
- Abowd, John M.** 2018c. "Tweeterial: formal privacy for social scientists." Twitter, November 30, 2018. [https://twitter.com/john\\_abowd/status/1068645579872497664](https://twitter.com/john_abowd/status/1068645579872497664).
- Abowd, John M.** 2019. "Staring Down the Database Reconstruction Theorem." Presentation, American Association for the Advancement of Science Annual Meeting, Washington, DC, February 16, 2019. <https://bpb-us-e1.wpmucdn.com/blogs.cornell.edu/dist/4/7616/files/2019/04/2019-02-16-Abowd-AAAS-Slides-Saturday-330-500-session-FINAL-as-delivered-liqsdg2.pdf>.
- Abowd, John M.** 2021. Declaration of John M. Abowd. Case no. 3:21-CV-211-RAH-ECM-KCN, US District Court for the Middle District of Alabama. April 13.
- Abowd, John M., Tamara Adams, Robert Ashmead, David Darais, Sourya Dey, Simson L. Garfinkel, Nathan Goldschlag et al.** 2023. "The 2010 Census Confidentiality Protections Failed, Here's How and Why." NBER Working Paper 31995.
- Abowd, John M., Gary L. Benedetto, Simson L. Garfinkel, Scot A. Dahl, Aref N. Dajani, Matthew Graham, Michael B. Hawes et al.** 2020. "The Modernization of Statistical Disclosure Limitation at the U.S. Census Bureau." <https://www2.census.gov/adrm/CED/Papers/CY20/2020-009-AbowdBenedettoGarfinkelDahletal-The%20modernization%20of.pdf>.
- Abowd, John M., and Michael B. Hawes.** 2023. "Confidentiality Protection in the 2020 US Census of Population and Housing." *Annual Review of Statistics and Its Application* 10: 119–44.
- Abowd, John M., Ian M. Schmutte, William N. Sexton, and Lars Vilhuber.** 2019. "Why the Economics Profession Cannot Cede the Discussion of Privacy Protection to Computer Scientists." Speech, Allied Social Science Associations Annual Meeting, Atlanta, GA, January 5. <https://ecommons.cornell.edu/handle/1813/60836>.
- Asquith, Brian, Brad Hershbein, Tracy Kugler, Shane Reed, Steven Ruggles, Jonathan Schroeder, Steve Yesiltepe, and David Van Riper.** 2022. "Assessing the Impact of Differential Privacy on Measures of Population and Racial Residential Segregation." *Harvard Data Science Review* (S2).
- Bambauer, Jane, Krishnamurthy Muralidhar and Rathindra Sarathy.** 2014. "Fool's Gold: An Illustrated Critique of Differential Privacy." *Vanderbilt Journal of Entertainment and Technology Law* 16 (4): 701–55.
- Barth-Jones, Daniel.** 2012. "The Debate over 'Re-identification' of Health Information: What Do We Risk?" *Health Affairs Forefront*. <https://doi.org/10.1377/forefront.20120810.021952>.
- Berman, John, Harriet McCombs, and Robert Boruch.** 1977. "Notes on the Contamination Method: Two Small Experiments in Assuring Confidentiality of Response." *Sociological Methods and Research* 6 (1): 45–63.
- Cantwell, Pat.** 2021. "How We Complete the Census When Households or Group Quarters Don't Respond." US Census Bureau, April 16 2021. <https://www.census.gov/newsroom/blogs/random-samplings/2021/04/imputation-when-households-or-group-quarters-dont-respond.html>.
- Capps, Kriston.** 2021. "Data Scientists Square Off Over Trust and Privacy in 2020 Census." *Bloomberg News*, August 12. <https://www.bloomberg.com/news/articles/2021-08-12/data-scientists-ask-can-we-trust-the-2020-census>.
- Commission on Evidence-Based Policymaking.** 2018. *The Promise of Evidence-Based Policymaking*. Washington, DC: Commission on Evidence-Based Policymaking.
- Committee on National Statistics.** 2020. *2020 Census Data Products: Data Needs and Privacy Considerations: Proceedings of a Workshop*. Washington, DC: National Academies Press.
- Committee on National Statistics.** 2023a. *Assessing the 2020 Census: Final Report*. Washington, DC: National Academies Press.
- Committee on National Statistics.** 2023b. *2020 Census Data Products: Demographic and Housing Characteristics File: Proceedings of a Workshop*. Washington, DC: National Academies Press.
- Daily, Donna.** 2022. "Disclosure Avoidance Protections for the American Community Survey." US Census Bureau, December 14. <https://www.census.gov/newsroom/blogs/random-samplings/2022/12/disclosure-avoidance-protections-acps.html>.
- Dajani, Aref N., Amy D. Lauger, Phyllis E. Singer, Daniel Kifer, Jerome P. Reiter, Ashwin Machanavajjhala, Simson L. Garfinkel et al.** 2017. "The Modernization of Statistical Disclosure Limitation at the U.S. Census Bureau." United Nations Economic Commission for Europe, Joint UNECE/Eurostat work session on statistical data confidentiality, Skopje, September 20–22. [https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2017/3\\_census\\_bureau.pdf](https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2017/3_census_bureau.pdf).
- Dick, Travis, Cynthia Dwork, Michael Kearns, Terrance Liu, Aaron Roth, Giuseppe Vietri, and Zhiwei Steven Wu.** 2023. "Confidence-Ranked Reconstruction of Census Microdata from Published



- Statistics.” *Proceedings of the National Academy of Sciences* 120 (8): e2218605120.
- Dillman, Don A., Eleanor Singer, Jon R. Clark, and James B. Treat.** 1996. “Effects of Benefits Appeals, Mandatory Appeals, and Variations in Statements of Confidentiality on Completion Rates for Census Questionnaires.” *Public Opinion Quarterly* 60: 376–89.
- Dinur, Irit, and Kobbi Nissim.** 2003. “Revealing Information While Preserving Privacy.” In *Proceedings of the Twenty-Second ACM Sigmod-Sigact-Sigart Symposium on Principles of Database Systems*, 202–10. New York: Association for Computing Machinery.
- Domingo-Ferrer, Josep, David Sánchez, and Alberto Blanco-Justicia.** 2021. “The Limits of Differential Privacy (And Its Misuse in Data Release and Machine Learning).” *Communications of the ACM* 64 (7): 33–35.
- Dwork, Cynthia.** 2011. “A Firm Foundation for Private Data Analysis.” *Communications of the ACM* 54 (1): 86–95.
- Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith.** 2006. “Calibrating Noise to Sensitivity in Private Data Analysis.” In *Theory of Cryptography*, edited by Shai Halevi, Tal Rabin, 265–84. Heidelberg: Springer Berlin.
- Foundations for Evidence-Based Policymaking Act of 2018.** P. L. No. 115–435. 132 Stat. 5529. (2019).
- Francis, Paul.** 2022. “A Note on the Misinterpretation of the US Census Re-identification Attack.” In *Privacy in Statistical Databases*, edited by Josep Domingo-Ferrer and Maryline Laurent, 299–311. Cham, Switzerland: Springer.
- Frey, James H.** 1986. “An Experiment with a Confidentiality Reminder in a Telephone Survey.” *Public Opinion Quarterly* 50 (2): 267–69.
- Garfinkel, Simson.** 2017. “Modernizing Disclosure Avoidance: Report on the 2020 Disclosure Avoidance Subsystem as Implemented for the 2018 End-to-End Test.” Presentation, Census Scientific Advisory Committee, Suitland, MD, September 15, 2017. <https://www2.census.gov/cac/sac/meetings/2017-09/garfinkel-modernizing-disclosure-avoidance.pdf>.
- Garfinkel, Simson.** 2023. “Comment to Muralidhar and Domingo-Ferrer (2023)—Legacy Statistical Disclosure Limitation Techniques Were Not an Option for the 2020 US Census of Population and Housing.” *Journal of Official Statistics* 39 (3): 399–410.
- Garfinkel, Simson L., John M. Abowd, and Sarah Powazek.** 2018. “Issues Encountered Deploying Differential Privacy.” In *WPES’18 Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, 133–37. New York: Association for Computing Machinery.
- Greenberg, Andy.** 2017. “How One of Apple’s Key Privacy Safeguards Falls Short.” *Wired Magazine*, September 15. <https://www.wired.com/story/apple-differential-privacy-shortcomings/>.
- Hansen, Mark.** 2018. “To Reduce Privacy Risks, the Census Plans to Report Less Accurate Data.” *New York Times*, December 5. <https://www.nytimes.com/2018/12/05/upshot/to-reduce-privacy-risks-the-census-plans-to-report-less-accurate-data.html>.
- Hauer, Mathew E., and Alexis R. Santos-Lozada.** 2021. “Differential Privacy in the 2020 Census Will Distort COVID-19 Rates.” *Socius* 7: 2378023121994014.
- Hawes, Michael.** 2022. “Reconstruction and Re-identification of the Demographic and Housing Characteristics File (DHC).” Presentation, Census Scientific Advisory Committee, March 17–18, 2022. <https://www2.census.gov/about/partners/cac/sac/meetings/2022-03/presentation-reconstruction-and-reidentification-of-the-dhc.pdf>.
- Hotz, V. Joseph, Christopher R. Bollinger, Tatiana Komarova, Charles F. Manski, Robert A. Moffitt, Denis Nekipelov, Aaron Sojourner, and Bruce D. Spencer.** 2022. “Balancing Data Privacy and Usability in the Federal Statistical System.” *Proceedings of the National Academy of Sciences* 119 (31): e2104906119.
- Jarmin, Ron.** 2019. “Census Bureau Adopts Cutting Edge Privacy Protections for 2020 Census.” US Census Bureau, February 15. [https://www.census.gov/newsroom/blogs/random-samplings/2019/02/census\\_bureau\\_adopts.html](https://www.census.gov/newsroom/blogs/random-samplings/2019/02/census_bureau_adopts.html).
- Jarmin, Ron S., John M. Abowd, Robert Ashmead, Ryan Cumings-Menon, Nathan Goldschlag, Michael B. Hawes, Sallie Ann Keller et al.** 2023. “An In-Depth Examination of Requirements for Disclosure Risk Assessment.” *Proceedings of the National Academy of Sciences* 120 (43): e2220558120.
- JASON.** 2022. *Consistency of Data Products and Formal Privacy Methods for the 2020 Census*. (JSR-21-02, January 11, 2022). The MITRE Corporation. <https://perma.cc/XJS8-ADX6>.
- Kenny, Christopher T., Shiro Kuriwaki, Cory McCartan, Evan T. R. Rosenman, Tyler Simko, and Kosuke Imai.** 2021. “The Use of Differential Privacy for Census Data and its Impact on Redistricting: The Case of the 2020 US Census.” *Science Advances* 7 (41): eabk3283.
- Kenny, Christopher T., Shiro Kuriwaki, Cory McCartan, Evan T. R. Rosenman, Tyler Simko, and Kosuke**

- Imai.** 2023. "Comment: The Essential Role of Policy Evaluation for the 2020 Census Disclosure Avoidance System." *Harvard Data Science Review* (S2).
- Lauger, Amy, Billy Wisniewski, and Laura McKenna.** 2014. "Disclosure Avoidance Techniques at the U.S. Census Bureau: Current Practices and Research." Research Report Series, Disclosure Avoidance #2014-02. Washington, DC: US Census Bureau.
- Lee, Jaewoo, and Chris Clifton.** 2011. "How Much Is Enough? Choosing  $\epsilon$  for Differential Privacy." In *International Conference on Information Security*, edited by Xuejia Lai, Jianying Zhou, and Hui Li, 325–40. Heidelberg: Springer Berlin.
- Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles.** 2022. "IPUMS National Historical Geographic Information System: Version 17.0 [dataset]." Minneapolis, MN: IPUMS. <http://doi.org/10.18128/D050.V17.0> (accessed June 3, 2023).
- McKenna, Laura.** 2018. "Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing." US Census Bureau Working Paper CES-18-47.
- McKenna, Laura.** 2019. "U.S. Census Bureau Reidentification Studies." US Census Bureau Working Paper CED-WP-2019-008.
- McKenna, Laura, and Matthew Haubach.** 2019. "Legacy Techniques and Current Research in Disclosure Avoidance at the US Census Bureau." US Census Bureau Working Paper CED-WP-2019-005.
- Menger, Elyse.** 2021. "Mermaids and Census Privacy Concerns." *Applied Geographic Solutions* September 2. <https://appliedgeographic.com/2021/09/mermaids-and-census-privacy-concerns/>.
- Mueller, J. Tom, and Alexis R. Santos-Lozada.** 2022. "The 2020 US Census Differential Privacy Method Introduces Disproportionate Discrepancies for Rural and Non-White Populations." *Population Research and Policy Review* 41 (4): 1417–30.
- Muralidhar, Krishnamurthy.** 2022. "A Re-examination of the Census Bureau Reconstruction and Reidentification Attack." In *Privacy in Statistical Databases*, edited by Josep Domingo-Ferrer and Maryline Laurent, 312–23. Cham, Switzerland: Springer.
- Muralidhar, Krishnamurthy, and Josep Domingo-Ferrer.** 2023a. "Database Reconstruction Is Not So Easy and Is Different from Reidentification." *Journal of Official Statistics* 39 (3): 381–98.
- Muralidhar, Krishnamurthy, and Josep Domingo-Ferrer.** 2023b. "A Rejoinder to Garfinkel (2023)—Legacy Statistical Disclosure Limitation Techniques for Protecting 2020 Decennial US Census: Still a Viable Option." *Journal of Official Statistics* 39 (3): 411–20.
- Muralidhar, Krishnamurthy, Josep Domingo-Ferrer, David Sanchez, and Steven Ruggles.** 2024. "Protecting Vulnerable Respondents: A Critical Analysis of the Privacy-Preserving Methods of the 2010 and 2020 Decennial Census." Unpublished.
- Ramachandran, Aditi, Lisa Singh, Edward Porter, and Frank Nagle.** 2012. "Exploring Re-Identification Risks in Public Domains." In *Tenth Annual International Conference on Privacy, Security and Trust*, 35–42. Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Rastogi, Sonya and Amy O'Hara.** 2012. *2010 Census Match Study: Final Report*. 2010 Census Planning Memoranda Series 247. Washington, DC: US Census Bureau.
- Reamer, Frederic G.** 1979. "Protecting Research Subjects and Unintended Consequences: The Effect of Guarantees of Confidentiality." *Public Opinion Quarterly* 43 (4): 497–506.
- Rodríguez, Rolando A.** 2021. "Disclosure Avoidance and the American Community Survey." Presentation, 2021 ACS Data Users Conference, May 20. <https://acsdatacommunity.prb.org/discussion-forum/m/2021-acs-conference-files/147/download>.
- Ruggles, Steven, Catherine Fitch, Diana Magnuson, and Jonathan Schroeder.** 2019. "Differential Privacy and Census Data: Implications for Social and Economic Research." *AEA Papers and Proceedings* 109: 403–08.
- Ruggles, Steven, and Diana L. Magnuson.** 2023. "'It's None of Their Damn Business': Privacy and Disclosure Control in the U.S. Census, 1790–2020." *Population and Development Review* 49 (3): 651–79.
- Ruggles, Steven, and David Van Riper.** 2022. "The Role of Chance in the Census Bureau Database Reconstruction Experiment." *Population Research and Policy Review* 41 (3): 781–88.
- Ruggles, Steven, and David Van Riper.** 2023. "Response to the Census Bureau Database Critique of 'The Role of Chance in the Census Bureau Database Reconstruction Experiment.'" Open Science Foundation Working Paper. <https://osf.io/my64r>.
- Sánchez, David, Josep Domingo-Ferrer, Krishnamurthy Muralidhar.** 2023. "Confidence-Ranked Reconstruction of Census Records from Aggregate Statistics Fails to Capture Privacy Risks and Reidentifiability" *Proceedings of the National Academy of Sciences* 120 (18): e2303890120.
- Santos-Lozada, Alexis R., Jeffrey T. Howard, and Ashton M. Verdery.** 2020. "How Differential Privacy

- Will Affect our Understanding of Health Disparities in the United States.” *Proceedings of the National Academy of Sciences* 117 (24): 13405–12.
- Schneider, Mike.** 2021. “People, Homes Vanish Due to 2020 Census’ New Privacy Method.” *Associated Press*, October 31. <https://apnews.com/article/religion-wisconsin-new-york-tampa-florida-68c96e7eb701da74ae7c8df3c3476705>.
- Singer, Eleanor, Hans-Jürgen Hippier, and Norbert Schwarz.** 1992. “Confidentiality Assurances in Surveys: Reassurance or Threat?” *International Journal of Public Opinion Research* 4 (3): 256–68.
- Sullivan, Teresa A., and Daniel L. Cork, eds.** 2022. *Understanding the Quality of the 2020 Census: Interim Report*. Washington, DC: National Academies Press.
- State of Alabama et al. v. US Department of Commerce et al.** 3:21-CV-211-RAH-ECM-KCN (2021).
- US Census Bureau.** 2021. “Census Bureau Sets Key Parameters to Protect Privacy in 2020 Census Results.” Press Release CB21-CN.42, June 9. <https://www.census.gov/newsroom/press-releases/2021/2020-census-key-parameters.html>.
- US Census Bureau.** 2022. “Privacy Loss Budget Allocation.” [https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/02-Demographic\\_and\\_Housing\\_Characteristics/2022-03-16\\_Summary\\_File/2022-03-16\\_Privacy-Loss\\_Budget\\_Allocations.pdf](https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/02-Demographic_and_Housing_Characteristics/2022-03-16_Summary_File/2022-03-16_Privacy-Loss_Budget_Allocations.pdf).
- US Census Bureau.** 2023a. “Detailed Demographic and Housing Characteristics File A (Detailed DHC-A) Proof of Concept.” [https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/03-Detailed\\_DHC-A/2023-01-31/Proof\\_of\\_Concept.pdf](https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/03-Detailed_DHC-A/2023-01-31/Proof_of_Concept.pdf).
- US Census Bureau.** 2023b. “2020 Census Data Table Guide.” <https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-table-guide-dhc-dp.xlsx>.
- Vadukul, Alex.** 2011. “Relaxing under Liberty’s Shadow.” *New York Times*, August 12. <https://www.nytimes.com/2011/08/14/nyregion/on-sundays-david-luchsinger-relaxes-under-lady-libertys-shadow.html>.
- Van Riper, David, Tracy Kugler, and Jonathan Schroeder.** 2020–2023. “IPUMS NHGIS Privacy-Protected 2010 Census Demonstration Data [database].” Minneapolis, MN: IPUMS. <https://www.nhgis.org/privacy-protected-2010-census-demonstration-data> (accessed June 3, 2023).
- Wagner, Deborah, and Mary Layne.** 2014. “The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications’ (CARRA) Record Linkage Software.” US Census Bureau CARRA Working Paper 2014-01.
- Wezerek, Gus, and Daniel Van Riper.** 2020. “Changes to the Census Could Make Small Towns Disappear.” *New York Times*, February 6. <https://www.nytimes.com/interactive/2020/02/06/opinion/census-algorithm-privacy.html>.
- Wines, Michael.** 2022. “The 2020 Census Suggests That People Live Underwater. There’s a Reason.” *New York Times*, April 21. <https://www.nytimes.com/2022/04/21/us/census-data-privacy-concerns.html>.
- Winkler, Richelle L., Jaclyn L. Butler, Katherine J. Curtis, and David Egan-Robertson.** 2022. “Differential Privacy and the Accuracy of County-Level Net Migration Estimates” *Population Research and Policy Review* 41 (2): 417–35.
- Zayatz, Laura, Jason Lucero, Paul Massell, and Asoka Ramanayake.** 2009. “Disclosure avoidance for Census 2010 and American Community Survey Five-Year Tabular Data Products.” Research Report Series 2009-10. Washington, DC: Statistical Research Division, US Census Bureau.



## **Gabriel Zucman: Winner of the 2023 Clark Medal**

Emmanuel Saez

**T**he 2023 John Bates Clark Medal of the American Economic Association was awarded to Gabriel Zucman, Associate Professor of Economics at the University of California, Berkeley, for his fundamental contributions to the study of inequality and taxation. Through meticulous empirical work and creative methodological approaches, Gabriel has revealed key trends about the concentration of global wealth, the size and distribution of tax evasion, and the tax-saving strategies of multinational companies. These findings have had a profound impact on the academic literature and on global policy debates. Methodologically, his work has contributed to reviving a centuries-long empirical tradition in the social sciences—pioneered long-ago by the likes of Gregory King and William Petty, and then in the twentieth century by Richard Stone and Simon Kuznets—that attempts to shed light on core economic and political issues through the creation of new measurement systems, which are sophisticated without being wedded to a unique theoretical perspective. In reviving this tradition, Gabriel has shifted the way contemporary economic research is done by showing that measurement can have a large impact in our field and on the world, inspiring many younger scholars to follow in his footsteps.

Gabriel Zucman was born in Paris. After his undergraduate studies—first in the “classe préparatoire B/L” of lycée Henri IV, then at École normale supérieure de Cachan—he chose to specialize in economics. In 2006, at barely 19, he founded a broad-audience review aimed at popularizing economic research, an early sign of

■ *Emmanuel Saez is Professor of Economics, University of California at Berkeley, Berkeley, California. His email address is [esaetz@berkeley.edu](mailto:esaetz@berkeley.edu).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.2.227>.



**Gabriel Zucman**

his entrepreneurial spirit. The review, which Gabriel served as director until 2013, invites scholars in economics and related disciplines to write articles summarizing their research and explaining how their work can illuminate the public debate of policy issues. The inaugural issue focused on taxation [1],<sup>1</sup> and the review still publishes in French, English, and Spanish today.<sup>2</sup>

Gabriel completed his master's degree at the Paris School of Economics in 2008. His dissertation, under the supervision of Thomas Piketty, investigated whether the wealthy fled the French wealth tax [2]. When he finished this work, the first signs of a devastating financial crisis were emerging. Academic economics seemed increasingly disconnected from the big issues of the time and, unsure about pursuing a PhD, he went to work at a financial firm. He started his position on the very day of the Lehman Brothers bankruptcy—September 15, 2008—tasked with explaining the global economic outlook to the firm's traders and clients in graphics-filled memos. This task was of course impossible (the severity and specificity of the crisis had caught most observers off guard) but it had the unexpected benefit of revealing to him the existence of a wealth of data that one could use to try to make sense of the world, including its dark corners that economists often do not like to talk about. He became fascinated by statistics on international capital

<sup>1</sup>We refer to Zucman's papers mentioned in this essay by number, as enumerated in Table 1.

<sup>2</sup>For the English-language text of *Regards croisés sur l'économie*, see <https://www.cairn-int.info/journal-regards-croises-sur-l-economie.htm>.

Table 1

## Cited Work by Gabriel Zucman

- 
1. Quelle fiscalité pour quels objectifs? *Regards croisés sur l'économie* 2007/1 (n° 1), initial issue.
  2. "Do High-Earners Flee the Wealth Tax? An Estimation Over the 1995–2006 Period." 2008. Paris School of Economics master thesis (in French).
  3. "The Missing Wealth of Nations: Are Europe and the U.S. net Debtors or net Creditors?" 2013. *Quarterly Journal of Economics* 128 (3): 1321–1364.
  4. "The End of Bank Secrecy? An Evaluation of the G20 Tax Haven Crackdown" (with Niels Johannesen). 2014. *American Economic Journal: Economic Policy* 6 (1): 65–91.
  5. "Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010" (with Thomas Piketty). 2014. *Quarterly Journal of Economics* 129 (3): 1255–1310.
  6. *La Richesse Cachée des Nations : Enquête sur les Paradis Fiscaux*. 2013. Le Seuil: Paris.
  7. *The Hidden Wealth of Nations: The Scourge of Tax Havens*. 2015. University of Chicago Press: Chicago.
  8. "Who owns the wealth in tax havens? Macro evidence and implications for global inequality." (with Annette Alstadsæter and Niels Johannesen). 2018. *Journal of Public Economics* 162: 89–100.
  9. "Tax Evasion and Inequality." (with Annette Alstadsæter and Niels Johannesen). 2019. *American Economic Review* 109 (6): 2073–2103.
  10. Tax Evasion at the Top of the Income Distribution: Theory and Evidence (with John Guyton, Patrick Langetieg, Daniel Reck and Max Risch). 2021. NBER working paper No. 28542.
  11. "Taxing Across Borders: Tracking Personal Wealth and Corporate Profits" 2014. *Journal of Economic Perspectives* 28 (4): 121–148.
  12. "The Missing Profits of Nations" (with Thomas Tørsløv and Ludvig Wier). 2023. *Review of Economic Studies* 90 (3): 1499–1534.
  13. "The Exorbitant Tax Privilege" (with Thomas Wright). 2018. NBER working paper No. 24983.
  14. "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data" (with Emmanuel Saez). 2016. *Quarterly Journal of Economics* 131 (2): 519–578.
  15. "Twenty Years and Counting: Thoughts about Measuring the Top Tail" (with Thomas Piketty and Emmanuel Saez). 2022. *Journal of Economic Inequality* 20: 255–64.
  16. "The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic Accounts" (with Emmanuel Saez). 2020. *Journal of Economic Perspectives* 34 (4): 3–26.
  17. "Trends in US Income and Wealth Inequality: Revising After the Revisionists" (with Emmanuel Saez). 2020. NBER working paper No. 27921.
  18. *The World Inequality Report 2022*. (with Lucas Chancel, Thomas Piketty, and Emmanuel Saez). 2022. Harvard University Press: Cambridge.
  19. "Distributional National Accounts: Methods and Estimates for the United States" (with Thomas Piketty and Emmanuel Saez). 2018. *Quarterly Journal of Economics* 133 (2): 553–609.
  20. "Simplified Distributional National Accounts" (with Thomas Piketty and Gabriel Zucman). 2019. *American Economic Association, Papers and Proceedings* 109: 289–295.
  21. *Global Tax Evasion Report* (with Annette Alstadsæter, Sarah Godar, and Panayiotis Nicolaidis). 2023. EU Tax Observatory report.
  22. *The Triumph of Injustice*. (with Emmanuel Saez). 2019. WW Norton: New York.
  23. "Revenue Effects of the Global Minimum Tax Under Pillar Two" (with Mona Baraké, Paul-Emmanuel Chouc, and Therese Neef). 2022. *Intertax* 50 (10): 689–710.
  24. "Progressive Wealth Taxation" (with Emmanuel Saez). 2019. *Brookings Papers on Economic Activity*, 437–511.
-

flows, in which one can see hundreds of billions of dollars flowing in and out of such places as the Cayman Islands and Luxembourg. He decided to go back to academia, pursue a PhD in economics, and explore the story behind these numbers.

His dissertation, completed in 2013, does not fit the mold of typical contemporary PhD theses in economics. It includes path-breaking work on the measurement of the wealth hidden in tax havens [3], the evaluation of policies implemented to curb this form of tax evasion [4], and (jointly with his PhD advisor Thomas Piketty) the construction of a centuries-long series of macroeconomic capital income and wealth for many countries [5] that would form the backbone of Piketty's (2014) best-seller *Capital in the 21st Century*. Perhaps most unusual for a young academic economist, Gabriel synthesized his thesis work on tax havens into a book published in France in 2013 [6], which was translated into 17 languages and remains his most cited solo work to date [7]. It is the first data-driven, scientific (and yet eminently accessible) treatment of the issue and it inspired the creation of a new field of studies on offshore wealth. The book illustrates some of the most distinct qualities of Gabriel's work: his ability to approach a complex topic with new eyes, to figure out what is important and missing in our understanding, and to illuminate what was previously unseen in a rigorous and yet engaging manner.

After his PhD, Gabriel was hired as Assistant Professor of Economics at the London School of Economics, before moving to the University of California Berkeley in 2015, where he obtained tenure in 2019. Since 2023, he also holds a position of Professor of Economics at the Paris School of Economics.

The American Economic Association is not the first to recognize Gabriel's accomplishments. He has received numerous prizes, including the Excellence Award in Global Economic Affairs from the Kiel Institute in 2017, the Best Young French Economist Prize awarded by *Le Monde* and *le Cercle des Economistes* in 2018, the Bernacer Prize and a Sloan Research Fellowship in 2019, and a Carnegie Fellowship in 2021.

Gabriel's Clark Medal citation says in part: "Through his entrepreneurial and creative pursuit of new data and methods for economic measurement, Gabriel Zucman has uncovered a range of fundamentally important facts quantifying the importance of tax evasion and measuring the rise of top income and wealth inequality." Two of his core topics of interest—tax evasion among the rich and measurement of top income and wealth—are inextricably linked. To measure top income and wealth properly, knowing how much is dissimulated by the rich and by the businesses they own is crucial. In this paper, we first provide some background on top-end inequality and tax evasion, and then discuss the pathbreaking contributions of Gabriel in each of these areas. Because being accessible has been central to Gabriel's approach, we will illustrate his accomplishments showing simple graphs drawn from his work that summarize his findings eloquently. Finally, we will discuss the policy impact that Gabriel's work has already had.



## Background

Perhaps the most common critique of capitalist economies is that while they foster economic growth, they also generate excessive inequality. The study of top levels of income and wealth is closely tied to tax statistics, because they are the best source of information to capture the rich—despite the long-held suspicion that the rich may not report all their income and wealth to the tax authority.<sup>3</sup> Over a century ago, Vilfredo Pareto (1896) discovered by looking at tabulations of income and wealth tax data for Swiss cantons that top tails of income and wealth distributions follow a power law, now named Pareto distributions. Fifty years later, Simon Kuznets (1953) estimated top income shares by dividing the income accruing to a top income group (for example, the top percentile) by total income economy wide estimated from the national accounts that he also helped to invent. These top income share statistics are more concrete than the abstract Pareto parameter of a power law—and hence have a greater impact in both the academic and policy debates. Kuznets documented the large decrease in US income concentration from 1913 to 1948, which formed the basis of his famous “Kuznets curve” theory that inequality would first rise and then fall with economic development. Lampman (1962), using estate tax data reweighted to represent the full population, documented a similar decline in US wealth concentration.

For decades afterwards, interest in using tax data to measure inequality waned as income and wealth concentration remained relatively stable and—by earlier historical standards—low, with the focus of distributional work shifting to data from the newly available microsurveys and focusing on the bottom rather than the top of the distribution.

Piketty (2001) and Piketty and Saez (2003) rekindled the study of the top tail using tax data by creating century-long top income shares for France and the United States. While income concentration had remained low in France, the United States experienced a sharp increase in income concentration starting precisely in 1980 after the neoliberal turn of the Reagan presidency. The striking contrast between France and the United States since 1980 showed that inequality trends could not be explained solely by technological progress, as posited by the earlier Kuznets (1953) explanations, or by skill-biased technological progress, as many US wage-inequality studies had proposed (as surveyed in Katz and Autor 1999).

Concerning the issue of US wealth inequality, the high-quality Survey of Consumer Finances showed significant increases in wealth concentration since the 1980s (Wolff 1995), as did the *Forbes* magazine list of the 400 richest Americans. However, estate tax data updating the Lampman (1962) study failed to validate such a surge in wealth concentration (Kopczuk and Saez 2004).

The main weakness of these earlier top income and wealth share studies was their reliance on income and wealth as reported on individual and estate tax

<sup>3</sup> This discussion is based partly on [15], in which one may find a more detailed exposition.

returns. Reported income may not include important components of income such as the undistributed profits of corporations that are particularly important at the top. Furthermore, tax evasion and tax avoidance may reduce reported incomes relative to real incomes and the extent of tax evasion and tax avoidance may also vary over time and across countries, depending on the strength of tax enforcement and the tax avoidance opportunities available. For the US economy, an enormous discrepancy between the booming wealth of the Forbes 400 and the stagnation of the largest estates of decedents suggested that growing tax avoidance/evasion could be an issue (Kopczuk and Saez 2004).

To be sure, an earlier literature on tax evasion and tax avoidance did exist, but it was rarely connected with the analysis of inequality. For the United States, studies of random audits by the Internal Revenue Service showed that adding estimates of evaded income to reported income slightly increases the concentration of income (Johns and Slemrod 2010). But the ability of the IRS to measure evaded income is limited, particularly for sophisticated forms of evasion used by the wealthy such as offshore tax evasion. While ample anecdotal evidence suggested that many rich people stashed wealth in tax havens to evade taxes and that tax havens had developed into a flourishing business model, there was no broadly applicable quantitative evidence about the extent of offshore tax evasion before Zucman's early work in this area [3].

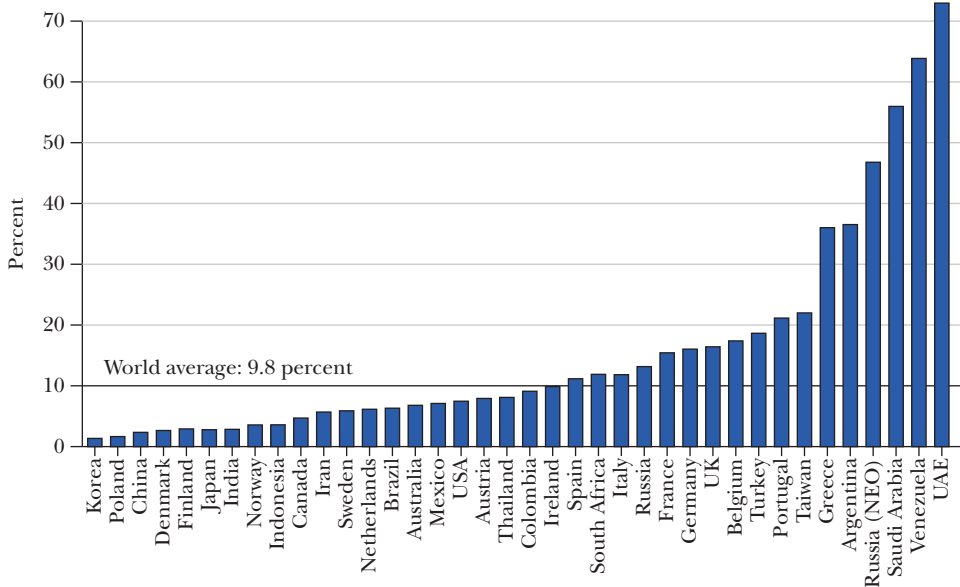
A somewhat larger literature in international business taxation studied the extent to which multinational companies avoid taxes by reporting profits in low-tax jurisdictions—oftentimes the same tax havens that cater to wealthy individuals. Because multinational companies report data on their foreign activities, a body of work suggested that such tax avoidance was significant (for example, Clausing 2009). Yet this literature was nearly impenetrable to the nonexpert because of the complexity of the tax avoidance schemes, the intricacies of the accounting firm data available and their lack of reliability, and perhaps as well an insider love for jargon and details, making it hard if not impossible for everybody else to see the big picture.

## **Offshore Tax Evasion**

In a series of papers and in his book *The Hidden Wealth of Nations* [3, 6, 7, 8, 9, 11], Gabriel developed methods to measure the wealth held in tax havens in a systematic manner, bringing much-needed quantitative assessments to this important issue. Gabriel measured both how much wealth is hidden in tax havens and how ownership of this wealth is distributed across the wealth distribution.

It is well-known that there is an anomaly in international statistics on securities (that is, stocks and bonds traded on markets): liabilities exceed assets. According to the statistics, the world as a whole is a net debtor (Lane and Milesi-Ferretti 2007). Gabriel's job market paper [3] notes that this discrepancy can be linked to household wealth assets being held in tax havens. To take a concrete example, if a French person owns US equities in her Swiss account, the US records a liability vis-à-vis the

Figure 1  
Offshore Wealth Relative to GDP in 2007, by Country



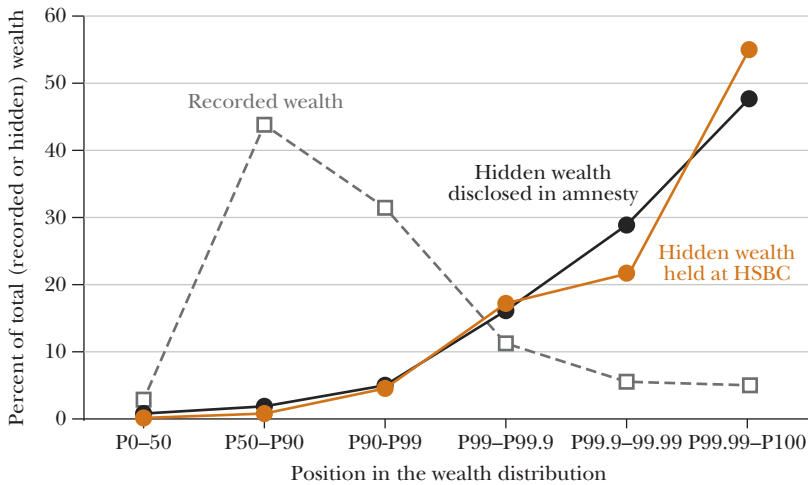
Source: [8], Figure 5.

Note: This figure depicts the amount of household wealth owned offshore by residents of each country as a percentage of GDP of the country, in 2007. The sample includes all the world's countries with more than \$200 billion in GDP in 2007. Offshore wealth is estimated by allocating the global offshore wealth estimated by [3], on the basis of the geographical distribution of bilateral cross-border bank deposits in offshore centers.

rest of the world, but France does not record any asset (as the French statistical authorities do not observe these claims), and neither does Switzerland (because these US equities do not belong to Swiss residents and are thus neither assets nor liabilities for Switzerland). Gabriel proposes a methodology to infer the size of global household offshore wealth from the pattern of anomalies seen in global investment data. This allows him to estimate that the equivalent of 8 percent of global household financial wealth—equivalent to about 10 percent of world GDP—is held in tax havens.

In [8], Gabriel exploits newly disclosed data by a number of prominent offshore financial centers—including Switzerland, Luxembourg, the Channel Islands, and Hong Kong—showing the amount of bank deposits that foreigners own in their banks by country of residence of foreigners. Using such data, Gabriel can allocate the global offshore wealth he found in [3] across countries. While on average, offshore wealth hidden in tax havens represents 10 percent of world GDP, he finds a great deal of heterogeneity across countries as he depicted on Figure 1. Scandinavian countries own the equivalent of only a few percent of GDP in offshore wealth,

Figure 2

**The Distribution of Recorded versus Hidden Wealth in Scandinavia**

Source: [9], Figure 4, Panel B.

Note: The figure shows the distribution of wealth in Scandinavia (Norway, Sweden, and Denmark), excluding offshore wealth (series recorded wealth), and the distribution of wealth held at the banking and financial services firm HSBC and disclosed by amnesty participants. Hidden wealth is dramatically more concentrated at the top than recorded wealth.

but this figure rises to about 15 percent in Continental Europe, and to as much as 60 percent in Russia, Gulf countries, and a number of Latin American countries.

While we can guess that most of the offshore wealth hidden abroad belongs to the wealthy, empirical work to quantify this effect was lacking. In [9], Gabriel made a path-breaking contribution in this direction using new microdata from recent leaks from offshore financial entities—the “Panama Papers” and HSBC “Swiss Leaks”—and from the results of tax amnesties, merged with administrative income and wealth tax records in Norway, Sweden, and Denmark. Offshore wealth from such leaks or reported through tax amnesties is incredibly concentrated at the very top of the distribution. In Figure 2, he depicted the distribution of wealth in Scandinavia (Norway, Sweden, and Denmark), excluding offshore wealth, and the distribution of wealth held at HSBC and disclosed by tax amnesty participants: about half of this offshore wealth belongs to the top 0.01 percent of the wealth distribution, about three-quarters belong to the top 0.1 percent, and over 90 percent belong to the top 1 percent. Since that work, a number of studies have found similar concentration of wealth in other contexts using leaks or amnesties (for an example using data from Colombia, see Londoño-Vélez and Ávila-Mahecha 2021). Therefore, it is plausible to use distributional estimates from Scandinavian countries produced in [9] and apply them to other countries. Using the aggregates by country of offshore wealth held by residents

discussed above, Gabriel was able to evaluate how wealth concentration changes when offshore wealth is included in [8]. Adding offshore wealth has only a modest impact for the United States, where wealth is already very concentrated and aggregate offshore wealth is fairly small: it increases the wealth share of the top 0.01 percent by less than 1 percentage point. However, the effects are much larger in many European countries where wealth is less concentrated and where aggregate offshore wealth is more important; for example, it increases wealth share of the top 0.01 percent in the United Kingdom from 2.7 to 4.4 percent (average in 2000–2009). The effect can be truly dramatic in countries with a lot of wealth hidden offshore: in Russia, the wealth share of the top 0.01 percent more than doubles from 5 to 12.5 percent.

This distributional analysis of offshore wealth has also shown that tax evasion at the top of the distribution is substantially higher than previously thought. The evidence from Scandinavia in [9] implies that the top 0.01 percent wealthiest Scandinavians evade about 25 percent of their taxes, in contrast with conventional audits that find evasion rates below 5 percent across the wealth distribution because they cannot capture offshore wealth. In [10], Gabriel applies those findings to the United States, showing that taking sophisticated tax evasion such as hidden wealth offshore into account, which cannot be detected by IRS random audit studies used to estimate tax evasion by income groups, dramatically shifts the picture of evasion across the income distribution. Instead of a fairly flat pattern of unreported income by income group from the IRS studies, unreported income as a fraction of true income rises from 7 percent in the bottom 50 percent of the income distribution to more than 20 percent in the top 1 percent. As a result, income that *should* be reported on US tax returns is more concentrated than income *actually* reported on tax returns.

## Multinational Tax Avoidance

A second strand of Gabriel's research quantifies the tax avoidance of multinationals using tax havens. Multinationals can use a variety of accounting strategies to maximize the share of their profits reported in tax havens, where their profits are taxed lightly if at all. In this way, the firms can minimize the share of their profits reported in higher tax countries, which are typically large and rich countries where the bulk of the real economic activity takes place. It is considered tax avoidance—as opposed to tax evasion discussed above—because these strategies are not outright fraud: they are devised by tax accountants and then vigorously defended when challenged by tax authorities. However, tax evasion and tax avoidance are closely related, because the line between aggressive tax avoidance that wins in courts and tax avoidance that loses and crosses into tax evasion territory is a thin one.

The important topic of tax avoidance by multinationals had been the subject of earlier academic research in the international tax literature, but the complexity of the institutional details made it hard going. In an early contribution, in what probably remains the simplest and most pedagogical contribution to understanding the big picture, Gabriel [11] constructs simple statistics illustrating the enormous rise of

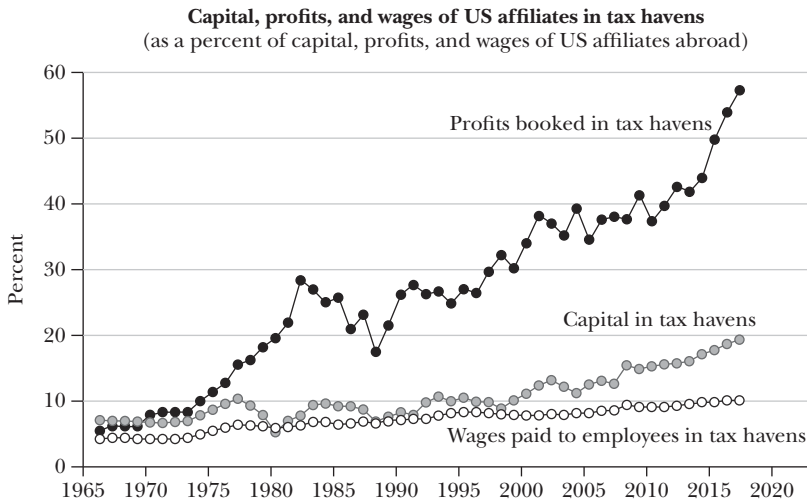
profit-shifting by US multinationals to tax havens: by the 2010s, more than half of the foreign profits reported by US multinationals are reported in tax havens. In [12], Gabriel mobilizes new macroeconomic data about global multinational companies and their foreign operations, known as “foreign affiliates statistics.” This step allows him to provide a first global and granular quantification of profit-shifting, with bilateral estimates of the amount of profit shifted out of any country A to any tax haven B. The analysis reveals that globally, 36 percent of the profits of multinationals reported abroad end up reported in tax havens and that US multinationals are particularly aggressive and shift almost twice as much as other multinationals (about 50 percent of their foreign profits, as opposed to about 30 percent for non-US multinationals). The paper provides a new comprehensive international database of profit-shifting available online at [MissingProfits.world](http://MissingProfits.world), and also provides revised versions of official statistics such as GDP and trade balances corrected for such profit-shifting.

Another important contribution of [12] is to show that multinational operations in tax havens appear an order of magnitude more profitable than local firms in the same tax havens, which strongly suggests that profits reported in tax havens do not represent real economic activity taking place in such tax havens, but rather inflated paper profits shifted for tax minimization purposes. Gabriel depicted a simple illustration of this phenomenon in Figure 3, which shows the evolution of the profits booked, tangible capital owned, and wages paid by US multinationals in tax havens since 1965, as a fraction of total foreign profits, capital, and wages of US multinationals (where “foreign” means outside the United States). The profit line shows that more than half of foreign profits of US multinationals are now booked in tax havens. However, only about 20 percent of tangible capital deployed abroad by US multinationals is deployed in tax havens. Moreover, only 10 percent of the foreign wage bill of US multinationals is for workers located in the tax havens. Therefore, real production including research and development, design services, or management, accounting, and legal services do not substantially move to tax havens because such moves would create a substantial wage bill in tax havens inconsistent with the data. Put simply, paper profits move to tax havens, while real economic activity in the form of real tangible capital and workers do so much less. This set of results have had a very large impact on the international tax literature that had focused on models of tax competition where countries use lower tax rates to compete for real economic activity, when the reality of tax competition is that countries compete to become the financial home of paper profits.

## **Measuring Top Income and Wealth**

The last strand of Gabriel’s research that I will discuss is the measurement of top income and wealth. In the end, measuring hidden wealth and shifted profits is done to achieve a better understanding of the wealth and income of the rich and the taxes they actually pay, which are issues of fundamental importance in any society.

Figure 3

**Paper Profits Moving to Tax Havens, Real Capital and Workers Less So**

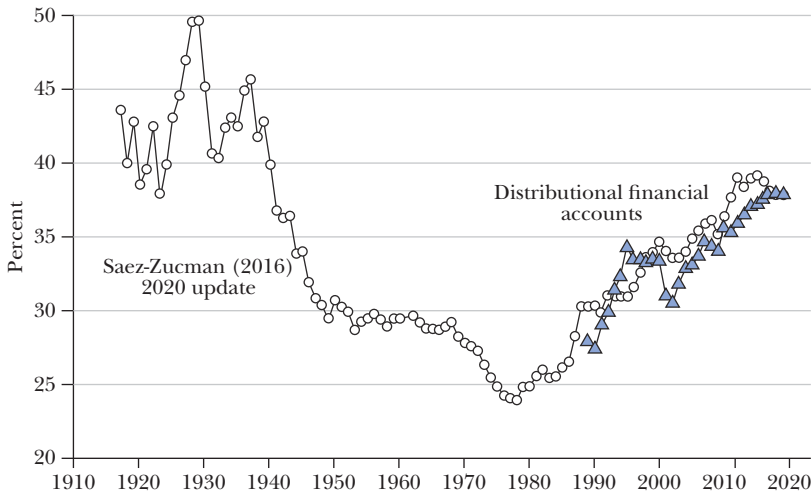
Source: Source: [13], Figure 3, Panel A.

Note: The figure depicts the evolution of the profits booked, tangible capital owned, and wages paid by US multinationals in tax havens since 1965, as a fraction of total foreign (that is, outside the United States) profits, capital, and wages of US multinationals.

**Wealth Inequality**

While many sources of information are available concerning US income inequality, much less is known about wealth. After all, the United States has no wealth tax (yet) that would provide such information systematically. Lists of the wealthy, such as the Forbes 400, along with data from the Survey of Consumer Finances told us that US wealth concentration had been increasing, while estate tax data told us it had not. Gabriel made progress on this question in [14] by applying another method to estimate wealth: the income capitalization method. We know a lot about capital income generated by wealth, because such income is generally taxable. The capitalization method seeks to infer wealth from capital income. In its simplest form, the method assumes that the rate of return on wealth is uniform within each asset class. Because the concentration of capital income flows such as dividends, capital gains, interest, or business profits have all increased sharply, the capitalization method finds that wealth has also increased sharply in recent decades. Figure 4 (reproduced from [16] appearing in this journal) depicts the wealth share of the top 1 percent in the US economy using the capitalization method since 1913 and compares it to the top 1 percent wealth share from the distributional financial accounts from the Federal Reserve Board based on the Survey of Consumer Finance data since 1989 (with small adjustments to make the Fed series directly comparable in terms of definition of

Figure 4

**Top 1 Percent Wealth Share in the United States**

Source: [16], Figure 1.

Note: This figure depicts wealth share held by the top 1 percent estimated using the capitalization method from [14] and the distributional financial accounts from the Federal Reserve Board using the same definition of wealth (market value of all financial assets net of all debts, excluding consumer durables and unfunded pensions).

wealth and family unit). Both series show a sharp increase in the wealth share of the top 1 percent wealth share, which increases from about 25 percent around 1980, a low level of wealth concentration by historical and international standards, to almost 40 percent in recent years, a very high level for an advanced economy in a democratic country. For a recent survey of wealth inequality over time and across countries, see [17]. Obviously, the capitalization method requires a lot of assumptions that have been debated in subsequent work leading to revised and better estimates [18]. Most notably, Smith, Zidar, and Zwick (2023) have used more granular internal tax data and alternative assumptions but find in the end remarkably similar results.

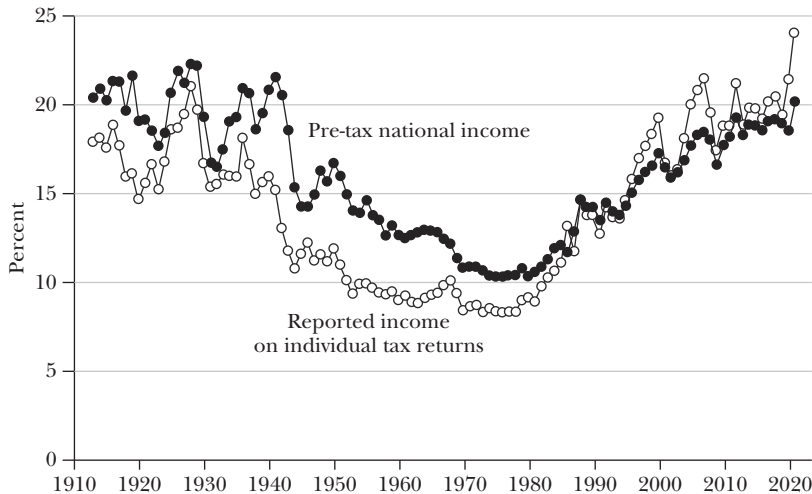
### Income Inequality

Gabriel has also contributed to the measurement of income inequality. Piketty and Saez (2003), using individual tax statistics, found that the concentration of reported income increased enormously since the 1970s, as depicted in Figure 5 using updated estimates to 2021—the top 1 percent reported income share more than doubled from 8 percent in the 1970s to over 20 percent in recent years.<sup>4</sup>

<sup>4</sup>The reported data series rank families by reported income excluding capital gains, but add back capital gains when computing the income share. This practice smooths out the lumpiness in capital gains realization, while taking into account this important source of income among the rich.



Figure 5

**Top 1 Percent Income Share: Reported Income versus National Income, 1913-2021**

Source: WID.world.

Note: This figure compares the share of reported income earned by the top 1 percent tax units in the United States (from Piketty and Saez 2003, updated; series including capital gains in income to compute shares but not to define ranks, to smooth the lumpiness of realized capital gains) to the share of pre-tax national income earned by the top 1 percent equal-split adults from [19].

Reported income concentration reached a record high in 2021, with a top 1 percent share of 24 percent as the strong recovery from the COVID-19 pandemic drove up asset prices and realized capital gains. However, reported income is about two-thirds of all national income earned by US residents. Undistributed corporate profits, fringe benefits of employees such as health insurance, and tax evasion are not included in reported income. Furthermore, the fraction of national income and product accounts reported in individual income tax data has declined from 70 percent in the late 1970s to about 60 percent in recent years [16]. The gap is even larger in survey data, such as the Current Population Survey, which do not capture top incomes well. Therefore, inequality measures using reported income or survey data are not consistent with macroeconomic measures of economic growth.

In [19], Gabriel and his co-authors pioneer “distributional national accounting,” which aims to allocate all national income across percentiles of the US income distribution. National income is conceptually the broadest definition of all economic income received by residents of the country. Starting from individual tax returns, all forms of income that are part of national income, but not included in reported individual income for tax purposes, are imputed and added. Figure 5 depicts the corresponding top 1 percent national income share on a pretax basis. The overall U-shape of income inequality over time remains, but slightly attenuated relative to the reported income series from Piketty and Saez (2003). Since

the 1970s, the top 1 percent still doubled from about 10 percent of the national income to about 20 percent in recent years. The modest effect of missing income on inequality is to be expected, giving that reported income still represents about two-thirds of national income. As explained in [20], it would require extreme, and hence highly unrealistic, equalization of unreported income to offset the visible and truly dramatic increase in reported income concentration depicted on Figure 5.

Distributional national income can be used to measure how macroeconomic growth is distributed across income groups. For the US economy, what stands out is the sharp change that took place since 1980. From the end of World War II to 1980, income growth was pretty much the same across each percentile of the income distribution, except somewhat smaller for the top 1 percent. Economic growth was truly lifting all boats; the macroeconomic growth rates told us how income all along the economic ladder was growing. Since 1980, however, economic growth has been skewed with hardly any income gains in real terms for the bottom 50 percent, solid income growth comparable to or better than the overall rate of growth only for the top 10 percent, and astonishing income growth for the top 1 percent. Distributional national accounts can also be used to consider income after taxes and adding transfers. The expansion of means-tested transfers has helped low-income groups, but the vast majority of this support is in-kind, particularly Medicaid and Medicare health insurance. Disposable cash income for the bottom 50 percent has not kept up with economic growth and has barely increased in real terms since 1980 (as shown in [16], Figure 6, in this journal). In recent years, the data show that the exceptionally generous transfers of 2020 and 2021 related to the COVID-19 pandemic increased the disposable incomes of the bottom 50 percent enormously, but only temporarily, with most of this extra support gone by early 2023 [21].

The distributional national account methodology has been applied to a wide range of countries [17]. It has produced series on wealth and income inequality that are also provided to the public in an accessible format via the World Inequality Database, available online, that Gabriel has helped create as one of the founding co-directors.

## **Policy Impacts**

It is hard to think of another economist whose academic research has had such a large impact on policy debates at such a relatively young age. Gabriel's research is tightly linked with several important policy developments in tax policy.

The problem of offshore tax evasion that Gabriel carefully documented is well on its way to being substantially reduced through systematic information reporting across countries, as advocated by Gabriel [6, 7] and, to be sure, many others. Following the impetus of the US Foreign Account Tax Compliance Act (FATCA) passed in 2010 under the Obama administration, more than 100 countries, including most tax havens, have now agreed to the Common Reporting Standard whereby their financial institutions will report income and wealth of their foreign clients to

the tax authorities of the home country of the clients. Gabriel, through the EU Tax Observatory he created thanks to funding from the European Commission, is monitoring these developments [21]. As data are generated, academic work will follow, a lot which will be done by young scholars inspired or directly supervised by Gabriel (see [21] for a recent overview of the early efforts). These developments would have felt utopian before 2010. They show that impediments to taxing the rich more fairly are not the inevitable result of a globalization, but rather the consequence of policy choices—and that these choices can be changed.

The problem of profit-shifting of US multinationals, also long seen as an intractable by-product of globalization, is being tackled through the proposed global 15 percent minimum tax on multinationals profits to which over 130 countries agreed to in October 2021. The Biden administration and US Treasury Secretary Janet Yellen played a leading role in spearheading this agreement, although ironically the United States has not yet enacted the global minimum tax itself. However, many countries, including the European Union member states, are moving forward. While the global minimum tax is modest in terms of the tax rate (15 percent) and generous in terms of exemptions, it shows global tax harmonization is not a pipe dream. Such a minimum tax had been advocated by Gabriel [22] and others. Before 2020 however, such a global agreement would have also been considered utopian, because the received wisdom was that globalization meant tax competition between countries to offer lower tax rates to multinationals. Future years will tell whether the global minimum tax can limit or reverse this trend. In [23], Gabriel has built a useful tool to provide revenue estimates for the currently proposed global minimum tax, as well as for alternative and more ambitious scenarios with higher tax rates and less generous exemptions.

The problem of growing wealth inequality in the United States and abroad has attracted many policy proposals, perhaps most famously a wealth tax on the ultra-rich proposed in the 2020 US Democratic presidential primaries by Elizabeth Warren and Bernie Sanders. Gabriel helped to shape and to estimate their budgetary effects [24]. While a US wealth tax is unlikely to happen anytime soon, the idea of taxing the huge wealth gains of the ultra-rich has become mainstream. For example, the Biden administration included such a proposal “the new billionaire minimum income tax” in its 2022 budget (White House 2022). A number of countries in the world are considering adding or reinitiating progressive wealth taxes as part of making their tax systems more progressive. Time will tell whether the revival of the idea of progressive wealth taxation will translate into legislation and whether these new wealth taxes can be well enforced and are indeed successful in improving the progressivity of our tax systems. Gabriel and the EU Tax Observatory will be watching.

## Conclusion

Because Gabriel’s work has been so innovative and relevant—indeed, it is at the center of many of the most hotly contested tax policy debates of the day—it has not come without pushback. Some have described it as painting an exaggerated

view of inequality and tax injustice. In reality—and I have witnessed the process firsthand—the data have radicalized Gabriel, rather than the other way around. The same is true for Thomas Piketty and for me: it is the increasing visibility of rising inequality and of its costs that has led us to consider more ambitious policy solutions.

But another form of radicalism is more specific to Gabriel. Piketty and I started our research careers with theoretical work that was well-anchored in the dominant neoclassical framework, such as optimal tax theory and political economy models, which gave us a lot of buy-in from the profession. In his formative years, Gabriel was more influenced by another intellectual tradition: that of the great British empirical social scientists of the last centuries—from Gregory King in the seventeenth and early eighteenth century, to Charles Booth in the late nineteenth and early twentieth century, to Richard Stone in the mid-twentieth century, and more recently Anthony Atkinson, who wanted to use economic statistics to understand (and reform) society—a tradition that flourished well before neoclassical economics. For the questions in which Gabriel was interested, he saw little need to encumber himself with the weight of the neoclassical apparatus. His methodological approach, reconnecting with that British tradition and applying it to the issues of the day (such as the arithmetic of international tax evasion and the rise of extreme wealth concentration) was radical and risky for a modern academic, but also fruitful: risky, because it drew the ire of economists who felt defensive towards this new empirical work that bypassed mainstream theorizing (or worse, showed that this theorizing could obscure reality rather than illuminate it); and fruitful, because it allowed Gabriel to see core problems with fresh eyes—and hence make genuine progress.

My colleague David Card, a former Clark Medalist, showed long ago that minimum wages do not always reduce employment (Card and Krueger 1994) and got pilloried for it by those who apparently could only reason in terms of the standard supply-and-demand competitive labor market where such a result is impossible. David Card was personally hurt by the experience enough that he vowed to never again work on minimum wages. Gabriel has followed a different strategy. He listened to his critics, and kept engaging with lengthy exchanges and incorporating valuable points [17]. His data series of offshore wealth, profit shifting, US inequality, and global wealth are all regularly updated,<sup>5</sup> incorporating new source data, refined methodologies, and lessons from the growing body of work on these issues—a truly unique approach in a profession focused on publishing papers rather than generating and updating data, and a testament to the seriousness of his approach. This path-breaking and meticulous work has shifted the way economic research is done by showing that bringing careful measurement to important but complex issues can have a large impact. It has already inspired many younger scholars to follow on his footsteps and will undoubtedly inspire many more.

<sup>5</sup>Gabriel's offshore wealth data series is at <https://atlas-offshore.world>. His data on profit-shifting is at <https://missingprofits.world>. His data on US inequality is at <https://gabriel-zucman.eu/usdina> and <https://realtimeinequality.org>. Finally, his data on global wealth is at <https://wid.world>.

## References

- Card, David, and Alan B. Krueger.** 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84 (4): 772–93.
- Clausing, Kimberly A.** 2009. "Multinational Firm Tax Avoidance and Tax Policy." *National Tax Journal* 62 (4): 703–25.
- Johns, Andrew, and Joel Slemrod.** 2010. "The Distribution of Income Tax Noncompliance." *National Tax Journal* 63 (3): 397–418.
- Katz, Lawrence F., and David H. Autor.** 1999. "Changes in the Wage Structure and Earnings Inequality." In *Handbook of Labor Economics*, Vol. 3A, edited by Orley C. Ashenfelter and David Card, 1463–1555. Amsterdam: Elsevier.
- Kopczuk, Wojciech, and Emmanuel Saez.** 2004. "Top Wealth Shares in the United States, 1916–2000: Evidence from Estate Tax Returns." *National Tax Journal* 57 (2): 445–87.
- Kuznets, Simon.** 1953. *Shares of Upper Income Groups in Income and Savings*. Cambridge, MA: NBER.
- Lampman, Robert J.** 1962. *The Share of Top Wealth-Holders in National Wealth, 1922–56*. Princeton: Princeton University Press.
- Lane, Philip R., and Gian Maria Milesi-Ferretti.** 2007. "The External Wealth of Nations Mark II: Revised and Extended Estimates of Foreign Assets and Liabilities, 1970–2004." *Journal of International Economics* 73 (2): 223–250.
- Londoño-Vélez, Juliana, and Javier Ávila-Mahecha.** 2021. "Enforcing Wealth Taxes in the Developing World: Quasi-experimental Evidence from Colombia." *American Economic Review: Insights* 3 (2): 131–48.
- Pareto, Vilfredo.** 1896. "La courbe de la répartition de la richesse." In *Ecrits sur la courbe de la répartition de la richesse*, edited by Giovanni Busino, 1–15. Geneva, Switzerland: Librairie Droz, 1965.
- Piketty, Thomas.** 2001. *Les hauts revenus en France au XXème siècle: Inégalités et redistributions 1901–1998*. Paris: Bernard Grasset.
- Piketty, Thomas.** 2014. *Capital in the Twenty-First Century*. Translated by Arthur Goldhammer. Cambridge, MA: Harvard University Press. Originally published as *Le Capital au XXIème siècle*. (Paris: Éditions du Seuil, 2013).
- Piketty, Thomas, and Emmanuel Saez.** 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118 (1): 1–39.
- Smith, Matthew, Owen Zidar, and Eric Zwick.** 2023. "Top Wealth in America: New Estimates under Heterogeneous Returns." *Quarterly Journal of Economics* 138 (1): 515–73.
- White House.** 2022. "President's Budget Rewards Work, Not Wealth with New Billionaire Minimum Income Tax." March 28, <https://www.whitehouse.gov/omb/briefing-room/2022/03/28/presidents-budget-rewards-work-not-wealth-with-new-billionaire-minimum-income-tax/>.
- Wolff, Edward.** 1995. *Top Heavy: The Increasing Inequality of Wealth in America and What Can Be Done About It*. New Press: New York.



## 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

Giancarlo Corsetti and Marco Buti survey “The first 25 years of the euro” (CEPR Policy Insight 126, February 2024, <https://cepr.org/publications/policy-insight-126-first-25-years-euro>). “At the age of 25, the euro has proven to be stable and resilient to existential threats. The great monetary experiment has gone through and passed a set of defining tests concerning its role in fostering the Single Market and providing an area of stability and inclusive growth for the residents of the union. . . . [T]he initial architecture and constitution of the euro was a ‘work in progress’ and required a dynamic development. . . . [L]eaving the euro area construction intentionally unfinished is dangerous and costly. The GFC [global financial crisis] and the ensuing sovereign debt crisis were arguably ‘too big to swallow’ for an incomplete union: an insufficient and incoherent response pushed the euro area

■ *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.38.2.245>.

into a painful existential crisis in 2010, much larger than in other regions of the world. The euro area may not survive another round of dismal performance caused by its own institutional and policy deficiencies. In the near future, climate change, the energy transition and geopolitical instability will bring more tail events to the plate of euro area policymakers . . .”

Albert Saiz discusses 30 different housing programs used in cities and countries around the world in “The Global Housing Affordability Crisis: Policy Options and Strategies” (IZA Policy Paper No. 203, October 2023, <https://docs.iza.org/pp203.pdf>). Here’s an excerpt from a case study: “Dutch Housing Associations (HA) are private, nonprofit enterprises that develop and manage affordable housing in the Netherlands. They account for approximately 75 percent of the three million rental homes and 35 percent of the entire housing stock, per 2016 estimates. HAs must lease 80 percent of their vacant units to low-income families and 10 percent to people with intermediate incomes. Ten percent can be leased to high-income families, which allows the associations to cross-subsidize their social mission. A government-regulated point system determines each unit’s rent, always substantially at below-market levels. Twenty-five percent of the total points are based on the tax-assessed market value of the property and 75 percent on the dwelling characteristics. The higher the number of points, the higher the allowed rental price. Points are also awarded based on factors such as size of the housing, facilities, and energy efficiency. The point system provides incentives to partially fund improvements with rental revenue growth. Subsequently, rents can only increase at a prespecified percentage annually (currently 3.3 percent). . . . HAs utilized a revolving fund model that—in addition to their equity—is sustained through rental revenue from tenants and sale proceeds from parts of their stock to investors. Excess funds are reinvested into renovating existing buildings, developing new affordable housing units, or developing neighborhood regeneration projects. . . . The associations do not require outside investors and can accept a lower or zero return on their equity. . . . Importantly, they do not utilize direct government subsidies. Instead, they benefit from cheap loans . . . In my view, the Dutch HAs model represents one of the most successful housing policies worldwide and is ripe for replication in other countries.”

Megan T. Stevenson offers skepticism about the policy implications of randomized controlled trials in “Cause, Effect, and the Structure of the Social World” (*Boston University Law Review*, December 2023, 103:7, pp. 2001–2027, <https://www.bu.edu/bulawreview/files/2023/12/STEVENSON.pdf>). “[M]ost reforms and interventions in the criminal legal space have little to no lasting effect when evaluated by RCTs [randomized controlled trials], and the occasional success usually fails to replicate when evaluated in other settings.” She backs this claim by discussing studies of counseling/therapy programs, criminal legal supervision, including intensive probation; scared-straight programs; work/job-training programs; drug testing, substance abuse counseling, and drug court; juvenile diversion; policing “hot spots”; and boot camps. “This Article shows that limited-scope, isolable interventions rarely lead to meaningful change. Those who desire meaningful change must therefore seek interventions outside the scope of what is evaluable via RCT. This includes changes



that are so multipronged and entangled that it is impossible to hold all else constant. This also includes changes that are so large in scope that experimental evaluation is infeasible. . . . It's hard to know what systemic reform will bring, not only because we cannot test its impact empirically, but because it's very hard to imagine a world that is otherwise the same as ours, while also being deeply, structurally different. When it comes to systemic reform, we are flying half-blind."

Gary Clyde Hufbauer and Megan Hogan argue that "America's payoff from engaging in world markets since 1950 was almost \$2.6 trillion in 2022" (Peterson Institute for International Economics, Policy Briefs 23-17, December 2023, <https://www.piie.com/publications/policy-briefs/americas-payoff-engaging-world-markets-1950-was-almost-26-trillion-2022>). Hufbauer and Hogan point to pre-2017 research on more than a dozen studies of international trade, which "calculated an average 'dollar ratio' of 0.24." They explain: "Simply put, the dollar ratio is the dollar increase in GDP divided by the dollar increase in two-way trade. In language familiar to economists, the dollar ratio is the elasticity of income (GDP) with respect to trade. Expressed another way, the calculation indicates that a 1 percent increase in trade yields a 0.24 percent increase in GDP—i.e., a \$1 billion increase in two-way trade increases GDP by \$240 million." They argue that more recent studies since 2017 suggest a higher "dollar ratio" of 0.30.

Anil Ari, Carlos Mulas-Granados, Victor Mylonas, Lev Ratnovski, and Wei Zhao describe historical and international experience since 1970 in "One Hundred Inflation Shocks: Seven Stylized Facts" (September 2023, IMF WP/23/190, <https://www.imf.org/en/Publications/WP/Issues/2023/09/13/One-Hundred-Inflation-Shocks-Seven-Stylized-Facts-539159>). Here are their seven facts. "Fact 1: Inflation is persistent, especially after a terms-of-trade shock. . . . Only in under 60 percent of episodes in the full sample (64 out of 111) was inflation resolved within 5 years after a shock. Even then, disinflation took on average over 3 years. . . . Fact 2. Most unresolved inflation episodes involved 'premature celebrations' . . . Fact 3: Countries that resolved inflation had tighter monetary policy . . . Fact 4. Countries that resolved inflation implemented restrictive policies more consistently over time . . . Fact 5. Countries that resolved inflation contained nominal exchange rate depreciation . . . Fact 6. Countries that resolved inflation had lower nominal wage growth . . . Fact 7. Countries that resolved inflation experienced lower growth in the short term but not over the 5-year horizon."

Davide Romelli discusses "Trends in central bank independence: a de-jure perspective" (BAFFI CAREFIN Centre Research Paper N. 217, February 2024, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4716704](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4716704); for a readable short overview, <https://cepr.org/voxeu/columns/recent-trends-central-bank-independence>). From the abstract: "This paper presents an extensive update to the Central Bank Independence—Extended (CBIE) index . . . extending its coverage for 155 countries from 1923 to 2023. The update reveals a continued global trend towards enhancing central bank independence, which holds across countries' income levels and indices of central bank independence. Despite the challenges which followed the 2008 Global financial crisis and the recent re-emergence of

political scrutiny on central banks following the COVID-19 pandemic, this paper finds no halt in the momentum of central bank reforms. I document a total of 370 reforms in central bank design from 1923 to 2023 and provide evidence of a resurgence in the commitment to central bank independence since 2016. These findings suggest that the slowdown in reforms witnessed post-2008 was a temporary phase, and that, despite increasing political pressures on central banks, central bank independence is still considered a cornerstone for effective economic policy-making.”

John Charles Bradbury, Dennis Coates, and Brad R. Humphreys review “The impact of professional sports franchises and venues on local economies: A comprehensive survey (*Journal of Economic Surveys*, September 2023, 1389–1431, <https://onlinelibrary.wiley.com/doi/10.1111/joes.12533>). “Between 1970 and 2020, state and local governments devoted \$33 billion in public funds to construct major-league sports venues in the United States and Canada, with the median public contribution covering 73% of venue construction costs. The prevalence of subsidized sports stadiums and arenas spawned an active economics literature evaluating their efficacy at stimulating economic activity. This literature contains near-universal consensus evidence that sports venues do not generate large positive effects on local economies. . . . Robust empirical findings documenting the impotence of professional sports in local economies likely reflect a simple theoretical explanation: consumer spending on sports represents a transfer from other local consumer spending, not net-new spending. Although sports games attract some nonlocals to spend money in the area, these visitors also crowd out other tourists attracted to other consumption amenities common to major US cities. Even with the presence outside visitors attracted by sports events, most consumer spending in and around pro sports venues derives from local residents; therefore, the opportunity cost of local sports consumption falls primarily on other competing local businesses, such as movie theaters, restaurants, and retail shopping. . . . Sports-related spending largely reflects a redistribution of existing spending by residents rather than increased local spending.”

## Distinguished Lectures

David A. Green delivered the Presidential Address to the Canadian Economic Association on the topic: “Basic income and the labour market: Labour supply, precarious work and technological change” (*Canadian Journal of Economics*, November 2023, pp. 1195–1220, <https://onlinelibrary.wiley.com/doi/abs/10.1111/caje.12698>; video of lecture at <https://www.economics.ca/cpages/presidential-address>). “I examine the question of whether a basic income could alter the quality of jobs and the level of wages by giving workers greater bargaining power. Working within a standard search and bargaining model, I show that there is no reason to think that it would have either of these effects. Theoretically, a basic income could even lead to decreases in wages—particularly at the low end. But more importantly, the analysis points to deficiencies in our models that become particularly evident if we consider policies in light of a goal of creating a more

just society. Once we look through that lens, the relationship of work to social respect and self-respect becomes more salient. The correct response to denigrating and otherwise problematic work conditions are policies that focus on elements of respect in the workplace. Treating those as reducible to monetary equivalents, as we tend to do in our models, misses the extent to which the work relationship is special and different from other exchanges in the economy. In addition, a focus on self-respect and social respect highlights people's need for community as the basis of that respect. Truly improving working conditions requires engaging the community of workers at a work site. Again, our models are not well suited to investigating how to do that. . . . But on the face of it, a basic income does not appear to be the best policy for improving job characteristics because it delivers only cash (which, again, is the wrong realm of consideration) and is given to individuals in the hope that they might use it to backstop community building behaviours without any particular evidence that it would actually do so."

Alan Blinder delivered the 2023 Daniel Patrick Moynihan Lecture in Social Science and Public Policy to the American Academy of Political and Social Science, on the topic "Economics and Politics: On Narrowing the Gap" (October 25, 2023, for text, see the Peterson Institute for International Economics website at <https://www.piie.com/commentary/speeches-papers/economics-and-politics-narrowing-gap>; for YouTube video, see <https://www.youtube.com/watch?v=jaF4Cjundy4>). "May I start by dispelling a myth? Perhaps because economists are frequently trotted out to support or oppose policies, perhaps because we have a Council of Economic Advisers right in the White House, perhaps because the powerful Federal Reserve—so ably represented here today—is dominated by economic thinking, many people believe that economists have enormous influence on public policy. In truth, apart from monetary policy, we don't. Almost a half century ago, George Stigler (1976, p. 351), later a Nobel prize winner, wrote that 'economists exert a minor and scarcely detectable influence on the societies in which they live.' Stigler was no doubt exaggerating to make his point. But he had a point. And things have not changed much since. . . . So here's my advice to economists interested in actual—as opposed to theoretical—policymaking. Don't forget about efficiency. It matters. We are right about that. But we may have to content ourselves with nibbling around the edges, below the political headline level, to make the details of a complex policy package less inefficient. Call it the theory of the third or fourth best."

## Symposia

*Cityscape*, published by the US Department of Housing and Urban Development, has a symposium on "100 Years of Federal Model Zoning" (2024, 25:3, <https://www.huduser.gov/portal/periodicals/cityscape/vol25num3/index.html>). The introductory essay, by Pamela M. Blumenthal, explains "It's Not Only Hoover's Fault: Reflections and Opportunities on the Centennial of the State Zoning Enabling Act": "In 1921, the U.S. Department of Commerce, under its then-Secretary

Herbert Hoover, supported the formation of an Advisory Committee on Zoning. The Advisory Committee's charge included aiding communities interested in the "promotion of the public welfare and the protection of property values . . . Zoning ordinances had been adopted in 8 cities by the end of 1916, another 68 cities by 1926, and an additional 1,246 municipalities by 1936, constituting 70 percent of the U.S. population." In another essay, Paul Cheshire offers "An International Perspective on the U.S. Zoning System." "In planning systems, the level to which decisions are rule-based, discretionary, or reflect local or wider societal interests varies globally. Internationally, the U.S. system is among the most locally controlled but significantly rule-based because of the use of zoning. In contrast, in the United Kingdom and a range of other countries, local politicians largely decide on development on a case-by-case basis. More local control and discretionary decisions increase the power of the 'not in my backyard,' or NIMBY, interest because development costs are highly localized, but benefits range over a wide area, even a whole country. This process tends to end with generally restricted development, resulting in higher housing and land costs. This problem is increasingly visible on both U.S. coasts. . . . The planning system common to Continental Europe, the Master Planning system, is more clearly rule-based, prescriptive, and detailed than the U.S. zoning system. Uses for every parcel are planned, and permission to develop is virtually automatic if the plan and any other relevant regulations are followed. In countries such as Germany, France, or the Netherlands, plan formulation and decision control has an important element, which is national, or at least regional."

*Social Philosophy and Policy* has published a twelve-paper symposium on topics of "Poverty, Agency, and Development." As one example, Johannes Haushofer and Daniel Salicath explore the evidence concerning "The psychology of poverty: Where do we stand?" (2024, 40:1, 150–184, <https://www.cambridge.org/core/journals/social-philosophy-and-policy/article/psychology-of-poverty-where-do-we-stand/0EE340EA852D6F3688C27F03024FA4DF>). "The purpose of this essay was to provide an overview of recent developments in the literature on the psychology of poverty. There has been significant progress in recent years, in particular, in establishing causality in the effect of income on psychological well-being; elucidating the precise functional form of psychological well-being with respect to income (satiation); and improving our understanding of the importance of relative income. Most saliently, the causal effect of income on psychological well-being is now robustly established. Research on the effects of scarcity and stress on economic decision-making has also made great strides in the past few years. However, the picture that emerges from these literatures is not as clear; individual studies are often statistically weak, provide conflicting evidence, and replication efforts have not always been successful. While the last word has perhaps not been spoken, in our view, the case for a poverty trap that operates through the effects of poverty on stress, decision-making, and cognition is currently not strong."

## Discussion Starters

Michael Bernick reminds us of “America’s Great Experiment With Jobs For The Underclass” (*Forbes*, February 21, 2024, <https://www.forbes.com/sites/michaelbernick/2024/02/21/americas-great-experiment-with-jobs-for-the-underclass>). “This month marks the fiftieth anniversary of the National Supported Work Demonstration project, America’s experiment to end the underclass through jobs. Supported Work is little remembered today. But it was one of the largest employment demonstrations undertaken by the federal government up to that time. . . . The Supported Work participants were drawn from four groups identified in the 1970s as the “hard to employ”: women who had been on welfare for at least 30 of the previous 36 months, ex-offenders, former drug addicts and unemployed out-of-school youth. The participants were provided with paid work experience, an array of training and social services, and placement assistance into regular jobs. . . . Supported Work operated from 1975 through 1979. During this time, around 10,000 workers were enrolled, split between the participant group and control group. . . . Only around a third of the 10,000 enrollees completed the program and went on to unsubsidized employment or to additional education/training. Further, the employment and income gains of the participants were not far above those of the control groups.”

Jan Feld, Corinna Lines, and Libby Ross provide evidence that even in academic research papers, “Writing matters” (*Journal of Economic Behavior and Organization*, January 2024, pp. 2378–397, <https://www.sciencedirect.com/science/article/pii/S0167268123004225>). From the abstract: “In this study, we estimate the effect of writing quality by comparing how 30 economists judge the quality of papers written by PhD students in economics. Each economist judged five papers in their original version and five different papers that had been language edited. No economist saw both versions of the same paper. Our results show that writing matters. Compared to the original versions, economists judge edited versions as higher quality; they are more likely to accept edited versions for a conference; and they believe that edited versions have a better chance of being accepted at a good journal.

Robert Francis provides some history at “Economic Ornithology: Before pesticides, birds were a farmer’s best defense against bugs. And the government’s economic ornithologists could tell you exactly how much each bird was worth” (Bird History substack, January 10, 2024, <https://birdhistory.substack.com/p/economic-ornithology>). “[The] US Department of Agriculture established the Section of Economic Ornithology in 1885. The following year it became the Division of Biological Survey, and was upgraded to the Bureau of Biological Survey in 1905. . . . In 1903, the Saturday Evening Post, for example, published a request that ‘every person in the United States who kills a bird is requested by the United States Government, not in a mandatory way, but as a matter of courtesy, to send the stomach and its contents to Washington.’ By 1916, the Bureau of Biological Survey had collected and analyzed the contents from more than 60,000 bird stomachs, which they used to determine whether each of the 400 species they studied was, on

balance, helpful or harmful to man. Researchers divided the stomach contents into 'good,' 'bad,' and 'neutral' categories, based on whether the partially-digested bug and plant matter was beneficial or harmful to farmers. . . . According to the Bureau of Biological Survey, native sparrows, who are 'specially efficient destroyers of weed seeds'; saved farmers \$35 million in 1906 by eating ragweed and crabgrass seeds. And during Nebraska's 1874 Rocky Mountain Locust infestation, a single Marsh Wren was calculated to have fed her brood of chicks enough grasshoppers to save \$1,743.97 worth of crops."



# MWI



## Meet the AEA Ombuds Team

The American Economic Association Ombuds Team is an **independent, impartial, confidential, and informal** third party resource to help AEA Members obtain information about AEA policies and procedures, discuss issues, and to consider options for resolving conflicts.

The Ombuds Team features three ombuds professionals with a range of experiences and specialties who can provide informal guidance on the evaluation of any interpersonal challenges including sexual harassment or discrimination issues, and determining how, when, or if a formal complaint to leadership may be an appropriate path or not.

### Contact your Ombuds to learn more:



**Nnena Odim**  
617-898-0938  
nnena-ombuds  
@mwi.org



**Charmhee Kim**  
845-213-6818  
charmhee-ombuds  
@mwi.org



**Larry Hunt**  
413-529-4035  
larry-ombuds  
@mwi.org

The ombuds team also provides regular educational sessions throughout the year on topics of interest to AEA members. For more details about upcoming workshops, please visit [www.aeaweb.org/ombuds](http://www.aeaweb.org/ombuds).



Learn more about the role of the MWI ombuds team resource and how it works to meet your needs.

# The Journal of Economic Perspectives: Proposal Guidelines



## Considerations for proposing topics and papers for JEP

Articles appearing in the journal are primarily solicited by the editors and associate editors. However, we do look at all unsolicited material. Due to the volume of submissions received, proposals that do not meet JEP's editorial criteria will receive only a brief reply. Proposals

that appear to have JEP potential receive more detailed feedback. Historically, about 10–15 percent of the articles appearing in our pages originate as unsolicited proposals.

Readers are also welcome to send e-mails suggesting topics for JEP articles and symposia and to propose authors for these topics. If the proposed topic is a good fit for JEP, the JEP editors will work to solicit paper(s) and author(s).

Correspondence regarding possible future articles for JEP may be sent (electronically please) to the assistant managing editor, Bradley Waldruff at [b.waldruff@aeapubs.org](mailto:b.waldruff@aeapubs.org). Papers and paper proposals should be sent as Word or pdf e-mail attachments.

### Philosophy and style

The *Journal of Economic Perspectives* attempts to fill part of the gap between refereed economics research journals and the popular press, while falling considerably closer to the former than the latter. **The focus of JEP articles should be on understanding the central economic ideas of a question, what is fundamentally at issue, why the question is particularly important, what the latest advances are, and what facets remain to be examined.** In every case, articles should argue for the author's point of view, explain how recent theoretical or empirical work has affected that view, and lay out the points of departure from other views.

We hope that most JEP articles will offer a kind of intellectual arbitrage that will be useful for every

economist. For many, the articles will present insights and issues from a specialty outside the readers' usual field of work. For specialists, the articles will lead to thoughts about the questions underlying their research, which directions have been most productive, and what the key questions are.

Articles in many other economics journals are addressed to the author's peers in a subspecialty; thus, they use tools and terminology of that specialty and presume that readers know the context and general direction of the inquiry. By contrast, **this journal is aimed at all economists, including those not conversant with recent work in the subspecialty of the author.** The goal is to have articles that can be read by 90 percent or more of the AEA membership, as opposed to articles that can only be mastered with abundant time and energy. Articles should be as complex as they need to be, but not more so. Moreover, the necessary complexity should be explained in terms appropriate to an audience presumed to have an understanding of economics generally, but not a specialized knowledge of the author's methods or previous work in this area.

The *Journal of Economic Perspectives* is intended to be scholarly without relying too heavily on mathematical notation or mathematical insights. In some cases, it will be appropriate for an author to offer a mathematical derivation of an economic relationship, but in most cases it will be more important that an author explain why a key formula makes sense and tie it to economic intuition, while leaving the actual derivation to another publication or to an appendix.

JEP does not publish book reviews or literature reviews. Highly mathematical papers, papers exploring issues specific to one non-US country (like the state of agriculture in Ukraine), and papers that address an economic subspecialty in a manner inaccessible to the general AEA membership are not appropriate for the *Journal of Economic Perspectives*. Our stock-in-trade is original, opinionated perspectives on economic topics that are grounded in frontier scholarship.



## Guidelines for preparing *JEP* proposals

Almost all *JEP* articles begin life as a two- or three-page proposal crafted by the authors. If there is already an existing paper, that paper can be sent to us as a proposal for *JEP*. However, given the low chances that an unsolicited manuscript will be published in *JEP*, no one should write an unsolicited manuscript intended for the pages of *JEP*. **Indeed, we prefer to receive article proposals rather than completed manuscripts.** The following features of a proposal seek to make the initial review process as productive as possible while minimizing the time burden on prospective authors:

- Outlines should begin with a paragraph or two that precisely states the main thesis of the paper.
- After that overview, an explicit outline structure (I., II., III.) is appreciated.
- The outline should lay out the expository or factual components of the paper and indicate what evidence, models, historical examples, and so on will be used to support the main points of the paper. The more specific this information, the better.
- The outline should provide a conclusion.
- Figures or tables that support the article's main points are often extremely helpful.

- The specifics of fonts, formatting, margins, and so forth do not matter at the proposal stage. (This applies for outlines and unsolicited manuscripts).
- Sample proposals for (subsequently) published *JEP* articles are available on request.
- For proposals and manuscripts whose main purpose is to present an original empirical result, please see the specific guidelines for such papers below.

The proposal provides the editors and author(s) an opportunity to preview the substance and flow of the article. For proposals that appear promising, the editors provide feedback on the substance, focus, and style of the proposed article.

After the editors and author(s) have reached agreement on the shape of the article (which may take one or more iterations), the author(s) are given several months to submit a completed first draft by an agreed date. This draft will receive detailed comments from the editors as well as a full set of suggested edits from *JEP*'s Managing Editor. Articles may undergo more than one round of comment and revision prior to publication.

## Guidelines for empirical papers submitted to *JEP*

*JEP* is not primarily an outlet for original, frontier empirical contributions; that's what refereed journals are for! Nevertheless, *JEP* occasionally publishes original empirical analyses that appear uniquely suited to the journal. In considering such proposals, the editors apply the following guidelines (in addition to considering the paper's overall suitability):

1. The paper's main topic and question must not already have found fertile soil in refereed journals. *JEP* can serve as a catalyst or incubator for the refereed literature, but it is not a competitor.
2. In addition to being intriguing, the empirical findings must suggest their own explanations. If the hallmark of a weak field journal paper is the juxtaposition of strong claims with weak

evidence, a *JEP* paper presenting new empirical findings will combine strong evidence with weak claims. The empirical findings must be robust and thought provoking, but their interpretation should not be portrayed as the definitive word on their subject.

3. The empirical work must meet high standards of transparency. *JEP* strives to only feature new empirical results that are apparent from a scatter plot or a simple table of means. Although *JEP* papers can occasionally include regressions, the main empirical inferences should not be regression-dependent. Findings that are not almost immediately self-evident in tabular or graphic form probably belong in a conventional refereed journal rather than in *JEP*.



# The AEA Departmental Seed Grants for Innovation in Diversity and Inclusion

## About the Awards

The American Economic Association and its Committee on the Status of Minority Groups in the Economic Profession (CSMGEP) have established an annual **AEA Departmental Seed Grants for Innovation and Diversity**. This one-time award is open to US-based economics departments in an amount up to \$5,000 to help establish a new “bridge program” or training program. Departments would use these funds to develop a program for underrepresented minorities (URM) such as persons who identify as American Indian, Alaskan Native, Black (not of Hispanic origin), Hispanic (including persons of Mexican, Puerto Rican, Cuban, and Central or South American origin), or Pacific Islander and other groups that are traditionally underrepresented in the field of economics.

For example, a department might create a mentoring program for URM graduate or undergraduate students, opportunities for URM students to do meaningful research assistant work, or a program that allows URM students who need additional preparation for graduate school to take a lighter class load in the first year or to take core economics courses over two years.

*Complete application criteria are available at the AEA website.*

## 2023 Award Winners

**Congratulations California State University, Bakersfield and Loyola Marymount University!**



CALIFORNIA STATE UNIVERSITY  
**BAKERSFIELD**



**Loyola Marymount  
University**

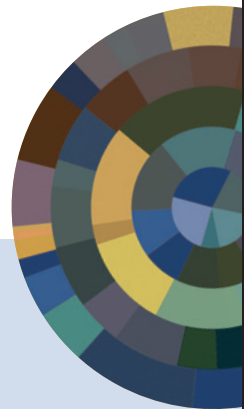
Previous award winners:

2022 **Florida International University**

2021 **Middlebury College and University of Kansas**

2020 **Georgia State University and Southern Methodist University**

[www.aeaweb.org/go/diversity-initiatives](http://www.aeaweb.org/go/diversity-initiatives)



# 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](mailto: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 © 2024 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](mailto:aeainfo@vanderbilt.edu).

## EXECUTIVE COMMITTEE

### Elected Officers and Members

#### *President*

JANET CURRIE, Princeton University

#### *President-elect*

LAWRENCE F. KATZ, Harvard University

#### *Vice Presidents*

NICHOLAS BLOOM, Stanford University

EBONYA L. WASHINGTON, Columbia University

#### *Members*

AMANDA BAYER, Swarthmore College

MELVIN STEPHENS, JR., University of Michigan

KEVIN LANG, Boston University

LISA M. LYNCH, Brandeis University

AMANDA E. KOWALSKI, University of Michigan

LARA D. SHORE-SHEPPARD, Williams College

#### *Ex Officio Members*

CHRISTINA ROMER, University of California, Berkeley

SUSAN C. ATHEY, Stanford University

### Appointed Members

#### *Editor, The American Economic Review*

ERZO F.P. LUTTMER, Dartmouth College

#### *Editor, The American Economic Review: Insights*

MATTHEW GENTZKOW, Stanford 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*

BENJAMIN OLKEN, Massachusetts Institute of Technology

#### *Editor, American Economic Journal: Economic Policy*

LUCAS W. DAVIS, University of California, Berkeley

#### *Editor, American Economic Journal: Macroeconomics*

AYŞEGÜL ŞAHİN, University of Texas at Austin

#### *Editor, American Economic Journal: Microeconomics*

NAVIN KARTIK, Columbia 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  
Nashville, TN

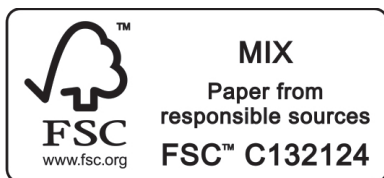
### ADMINISTRATORS

#### *Director of Finance and Administration*

BARBARA H. FISER

#### *Convention Manager*

REBEKAH LOFTIS



*The Journal of*  
***Economic Perspectives***

---

Spring 2024, Volume 38, Number 2

---

**Symposia**

***How Research Informs Policy Analysis***

**Staff of the Congressional Budget Office**, “How Economists Could Help Inform Economic and Budget Analysis Used by the US Congress”

**Cass R. Sunstein**, “The Economic Constitution of the United States”

**Wendy Edelberg and Greg Feldberg**,

“The Financial Crisis Inquiry Commission and Economic Research”

**Emily Oehlsen**, “Philanthropic Cause Prioritization”

***Labor Market and Macroeconomics***

**Gadi Barlevy, R. Jason Faberman, Bart Hobijn, and Ayşegül Şahin**,  
“The Shifting Reasons for Beveridge Curve Shifts”

**Loukas Karabarbounis**, “Perspectives on the Labor Share”

**Richard Rogerson**, “Why Labor Supply Matters for Macroeconomics”

**Marianna Kudlyak**, “How Cyclical Is the User Cost of Labor?”

***Privacy Protection and Government Data***

**Claire McKay Bowen**, “Government Data of the People, by the People,  
for the People: Navigating Citizen Privacy Concerns”

**Steven Ruggles**, “When Privacy Protection Goes Wrong:  
How and Why the 2020 Census Confidentiality Program Failed”

**Article**

**Emmanuel Saez**, “Gabriel Zucman: Winner of the 2023 Clark Medal”

**Features**

**Timothy Taylor**, “Recommendations for Further Reading”

