MAINTAINING THE STRENGTH OF AMERICAN CAPITALISM

Foreword by HENRY M. PAULSON, JR. and ERSKINE BOWLES

Edited by MELISSA S. KEARNEY and AMY GANZ
MAINTAINING THE STRENGTH OF AMERICAN CAPITALISM

Foreword by HENRY M. PAULSON, JR. and ERSKINE BOWLES

Edited by MELISSA S. KEARNEY and AMY GANZ

DECEMBER 2019
We thank the many individuals who were instrumental in the production of this volume. First, the members of the Economic Strategy Group, whose thoughtful suggestions and discussion motivated the subject of this book. We are grateful to the authors of this volume, whose contributions advanced our understanding of complex topics and challenged us to think more deeply about potential policy responses.

We are grateful for the individuals and organizations whose generous support makes possible the work of the Economic Strategy Group. We are grateful to the leadership of The Aspen Institute for its continued support of the Economic Strategy Group’s mission. We thank the staff who invested countless hours in producing this volume: Andy Morimoto and Kelley Folino for their helpful feedback; Dwyer Gunn and Kate Wheeling for excellent proofreading and editing assistance; Emily Vincent and Anne Hawkins for superb project management; Daniel Schiff for insightful research assistance; and Krista Martin, Sogand Sepassi, and Steven Johnson for their thorough editing and layout work.

Henry M. Paulson, Jr.  Erskine Bowles  
Co-Chair  Co-Chair

Melissa S. Kearney  Amy Ganz  
Director  Deputy Director

This volume was produced to clarify different perspectives on some of the current challenges confronting American capitalism. Authors are invited to share their views about complex policy issues regardless of whether or not the co-chairs, staff, or members of the Economic Strategy Group agree with them. Therefore, the views expressed herein are those of the authors and do not necessarily reflect the views of the ESG members or the organizations they represent.
# Table of Contents

**Foreword** ................................................................. 1
Erskine Bowles and Henry M. Paulson, Jr.

**Introduction** ............................................................ 3
Melissa S. Kearney and Amy Ganz

**Part I: Market Concentration**

**Causes, Consequences, and Policy Responses to Market Concentration**  .................................................. 14
Thomas Philippon

**Concerns About Concentration** .................................... 34
Nancy L. Rose

**Part II: Rising Federal Debt and Slowing Economic Productivity**

**Fiscal Policy With High Debt and Low Interest Rates**  ................................................................. 78
William Gale

**Can Innovation Policy Restore Inclusive Prosperity in America?** .................................................. 116
John Van Reenen

**Part III: Increasing Government Redistribution in Response to Income Inequality and Declining Economic Mobility**

**The Economics of Medicare for All**  ................................................. 136
Craig Garthwaite

**Universal Basic Income (UBI) as a Policy Response to Current Challenges** ........................................ 162
Melissa S. Kearney and Magne Mogstad
Wealth Taxation: An Overview of the Issues .................................................. 180
Alan D. Viard

Policy Options for Taxing the Rich ................................................................. 200
Lily Batchelder and David Kamin

Author Biographies ....................................................................................... 235
Foreword

By Erskine Bowles and Henry M. Paulson, Jr.

The American economic system has always been the foundation of our national strength. But this foundation is showing cracks—from high levels of income inequality, declining economic mobility, and persistent economic insecurity among low- and middle-income Americans.

Many now conclude that our economic system is broken. Recent polling data show that trust in capitalism is declining, especially among younger people. A 2018 Gallup poll found that less than half of respondents (45%) ages 18-29 held positive views of capitalism. This shift represents a 20-point decline since 2010 in the share of young adults’ who held positive views of capitalism.

The upshot is clear: American capitalism is in trouble. We need to strengthen our system to ensure that more people participate in our economic success. This means updating and adjusting our policies to ensure the outcomes of our market-based economy are consistent with fundamental American values of freedom, opportunity, and equality.

Doing so isn’t just an imperative for economic reasons. We believe that strengthening capitalism is as important for the health of the American economy as it is for the strength of our democracy. High levels of economic inequality will only contribute to increasing political dysfunction.

The essays contained in this volume seek to clarify the lines of debate on some of the greatest economic policy challenges of our time and present evidence-based analysis on how to address them. It examines the hypothesis that growing market concentration is inhibiting a dynamic and competitive economy. Next, it examines the health of America’s fiscal situation and what it implies about the continued strength of our market-based economy. Finally, it takes a hard look at recent policy proposals that would dramatically raise taxes on the rich and expand access to public benefit programs in response to high levels of income inequality and declining economic mobility.

The perspectives presented in this volume are not intended to represent the consensus view of Aspen Economic Strategy Group members. Our goal is to equip policymakers with the best analysis available to better inform decision making and to help Americans better understand the difficult trade-offs our leaders face in making such decisions.
There is no single solution to the challenges facing the American economy. The important role of evidence-based policies with bipartisan appeal, however, is difficult to overstate. This volume cannot claim to represent the end of thinking on ways to strengthen American capitalism, but we believe it provides a useful start.
Introduction

By Melissa S. Kearney and Amy Ganz

A national debate about the strength and fairness of American capitalism is taking place against a backdrop of vast levels of income and wealth inequality, growing pessimism about the state of economic opportunity and mobility, increased market concentration in many sectors, and a precarious fiscal situation. Restoring the promise of America’s capitalist system will require policies that enable more Americans to succeed in our market-based economy. Designing effective policies requires an accurate diagnosis of what is ailing American capitalism in order to effectively strengthen it. This volume brings to bear perspectives from leading subject matter experts on critical issues.

This book is organized around three broad economic challenges facing the United States. Section I addresses the widespread concern that increasing market concentration in many sectors is stifling competition and undermining a more dynamic economy. Section II explores the federal government’s unsustainable deficit and debt position and the associated concern that such trajectories imperil the long-run stability and security of the American economy. Section III considers a range of current policy ideas—including a federal wealth tax, “Medicare for All,” and universal basic income—that would dramatically change our economic institutions and policies in order to achieve greater progressivity through taxes and government spending.

Part I: Market Concentration

Competition in product and labor markets allows for the efficient allocation of resources and is the foundation of the market economy. Competition drives innovation, lowers prices, and increases output. Yet, there is reason to worry that competition in the U.S. product and labor markets is dwindling. Fewer new firms are entering some markets; aggregate measures of firm concentration are rising; and mergers and acquisitions have increased. These trends coincide with a falling labor share of income and rising profits among market-leading firms, stoking concern among many observers about increasing firm monopoly power over consumer prices and monopsony power over employee wages.

However, there is not consensus among scholars as to what the trends in market concentration are, let alone what they imply for consumer and worker welfare. A recent academic paper shows that concentration in local markets has actually
declined, even as aggregate measures have increased.\textsuperscript{1} Furthermore, to the extent that some industries have experienced increased concentration and a decline in the labor share of profits, it is not clear that the trends reflect anticompetitive forces or an erosion of labor institutions. For instance, a recent paper presents evidence that the decline in the labor share is consistent with a rise of “superstar firms.” In this framework, changes in the economic environment that advantages the most productive firms in an industry lead to an increase in product market concentration and a fall in labor share. Competing explanations for recent trends in market concentration have very different policy implications.\textsuperscript{2}

This volume features two chapters on the state of market concentration and competition in the United States, and potential policy responses to these trends. Nancy Rose, the Charles P. Kindleberger Professor of Applied Economics and Head of the Department of Economics at the Massachusetts Institute of Technology, encourages caution in interpreting aggregate trends in market concentration as signifying meaningful reductions in competition. Her memo highlights substantial measurement challenges that call into question the interpretation of these measured trends.

Despite Rose’s skepticism that the documented increases in market concentration actually reflect reduced market competition, her assessment of the relevant economic and legal landscape lead her to suggest reforms that would promote greater market competition and enhance antitrust enforcement from the Department of Justice and other federal agencies.

Economist Thomas Philippon, Professor of Finance at the Stern School of Business at New York University, takes a more aggressive stance on the issue of whether market concentration is impeding competition. He argues that although increasing market concentration is not always harmful to the economy, rising concentration since the early 2000s has produced market inefficiencies and reduced investment and productivity growth. He also attributes much of the increase in market concentration since the early 2000s to weak antitrust enforcement and new barriers to entry.

In response, Philippon proposes regulatory reform at the federal and state level that would promote greater competition, including more vigorous antitrust enforcement and occupational licensing reform. Philippon also considers the state of digital competition, an area that has garnered much interest and concern in recent months. There, he encourages strengthening regulations that would promote consumer welfare by increasing platform interoperability and data portability.


Part II. Rising Federal Debt and Slowing Economic Productivity

2.1 U.S. Fiscal Debt

According to recent estimates from the Congressional Budget Office, the U.S. debt as a share of the economy is on track to surpass its previous World War II high of 106% of Gross Domestic Product (GDP) as soon as 2037 and climb to 144% by 2049 (see Figure 1). The projected surge in long-term debt is driven by a combination of increased entitlement spending on an aging population and interest payments on the national debt. Government spending as a share of GDP is on track to increase from 20.7% today to 28.2% by 2049. In contrast, revenues under current law are expected to remain relatively flat, increasing from only 16.5% to 19.5% of GDP over the next 30 years.³

Prior beliefs about fiscal constraints are being challenged by historically low interest rates and unprecedented levels of government debt. Current projections show that the United States’ debt and deficit trajectory will continue to rise indefinitely as a share of the economy absent policy change. Yet, long-term interest rates on government debt remains historically low (Figure 2). Moreover, many mainstream, macroeconomic models imply interest rates are expected to remain well below historical levels for the foreseeable future.

To what extent should policymakers be concerned about stabilizing the fiscal trajectory now as opposed to waiting until later? In his chapter, William Gale, the Arjay and Frances Miller Chair in Federal Economic Policy at the Brookings Institution, argues that the federal fiscal outlook is unsustainable even if interest rates remain below the growth rate for the next 30 years. He argues that short-term policy responses should focus on investments that are tax-financed rather than debt-financed, and that, most importantly, policymakers should enact a debt reduction plan that is gradually implemented over the medium to long term. He also considers the merit of enacting various fiscal targets, such as a benchmark of debt-to-GDP or interest payments as a share of the economy.

### 2.2 Technological Innovation

Any proposal for new government spending or changes to the tax code will raise important questions about what is *fiscally feasible* in light of the federal government’s debt and deficit trajectory. Of course, not all spending proposals will have the same long-run impact on the budget, dollar-for-dollar, since some expenditures will enhance economic growth and eventually create more government revenue.

Long-run economic growth is driven by productivity gains and, in particular, technological innovation. Government investments in technological innovation have played a significant role in America’s economic prosperity. But, will America remain at the forefront of technological advancement, especially in light of the massive government investments in new technology in China and other East Asian countries?
In his chapter, MIT economist John Van Reenen argues that the private sector alone does not invest in developing new technologies at a socially optimal level. Further, he argues that large, positive economic spillovers exist from government-sponsored basic research and private-sector technological innovation, which justify far greater public investment in such activities. Van Reenen considers the merits of a U.S. industrial policy that would be comprised of tax credits, direct subsidies, and human capital investments, all of which have been shown to produce net positive research and development activity in the public and private sectors.

**Part III. Increasing Government Redistribution in Response to Income Inequality and Declining Economic Mobility**

### 3.1 Rising Income Inequality and Declining Economic Mobility

Trends showing widening income inequality, stagnant median income growth, and declining economic mobility suggest economic opportunity is declining and economic outcomes are disappointing for many Americans. The average household in the bottom 20% experienced annual income gains of 0.8% per year between 1979 and 2015, while average incomes in the top 1% grew by 3.4% per year over the same period. Figure 3 illustrates overall income growth by household income quintile since 1979. After taxes and transfers, income growth for the average household in the bottom quintile from 1979 to 2015 approached 80% (1.6% per year), with about half of that growth composed of means-tested transfers (i.e., Medicaid, Children’s Health Insurance Program (CHIP), Supplemental Nutrition Assistance Program (SNAP), Supplemental Security Income (SSI)). Meanwhile, those in the middle 60% of the distribution saw income gains of 45% to 51% (1% to 1.2% per year), post taxes and transfers. The distribution of post-transfer income growth by quintile is illustrated in Figure 4.

While income inequality has increased over time, intergenerational mobility—the share of adults who eventually earn more than their parents—has decreased. Data reveals that today’s generation of adults are less likely to have surpassed their parents’ level of income than the generation before them. Figure 5 shows that the share of children who surpass their parents’ earnings as adults has declined sharply over the past few decades. Only 50% of individuals born in the 1980s earn more than their parents did at their age, as compared to more than 90% of adults born in the 1940s.

---


Low rates of economic mobility result from both lower economic growth and increasing income inequality. Had economic growth maintained the levels the United States experienced in the mid-20th century—and assuming no changes to the current distribution of growth—62% of individuals born in 1980 would earn more than their parents did. In contrast, assuming current growth rates but allowing
income to be distributed as it was in the 1940s, 80% of children would exceed their parents’ income level.\(^6\)

**Figure 5. Absolute Economic Mobility Has Declined**

---

### 3.2 Policy Responses

Many Americans report favoring bold policy proposals to reduce income inequality and foster greater economic mobility. A number of leaders are calling for a dramatic expansion of government programs and benefits to deliver more widespread economic security. Proposals for universal benefits, including for childcare, health care, post-secondary education, and even basic income, abound in today’s policy debates, each of which would expand the scope and reach of the social safety net higher up in the income distribution.

In his chapter, Craig Garthwaite, the Herman R. Smith Research Professor in Hospital and Health Service at the Kellogg School of Management, describes the economic trade-offs of expanding the government’s role in financing and regulating health care through a single-payer system such as Medicare for All. Such a policy would expand access and potentially reduce the costs of health care for beneficiaries. Garthwaite highlights the trade-offs and downstream consequences of such a significant structural change to the U.S. health-care system, including potential changes in the quality and quantity of medical services supplied and the willingness of drug innovators to invest in the research and development (R&D) of new products. Garthwaite also highlights several alternative policy approaches that could promote the dual goals of increasing access to and affordability of health care.

---

6 Chetty et al., 2017.
Another policy idea that has enjoyed renewed popularity in recent years is the idea of a government Universal Basic Income (UBI) guarantee. Aspen Economic Strategy Group (AESG) members Melissa Kearney and Magne Mogstad argue that any substantial UBI program would be extremely expensive—potentially costing up to $3 trillion per year, or roughly three-quarters of the entire U.S. federal government budget—and yet would do little to address the underlying challenges driving increased income inequality and decreased economic mobility. They argue that a UBI would spend a great deal of public resources providing income assistance to individuals who don’t need it. Furthermore, they argue, a UBI could exacerbate the very challenges the policy is intended to address by discouraging work and diverting resources away from existing benefit programs that promote human capital development and economic mobility.

Recent polling data show that a large majority of registered voters believe the tax system favors the wealthy (73%) and upper-income individuals should pay more in taxes (63%). At the time of this writing, two presidential candidates and numerous lawmakers in Congress have responded by putting forward proposals that would raise dramatically more tax revenues from the wealthiest Americans, including by taxing wealth directly.

This volume features two chapters written by leading tax experts on the issues of wealth taxation and income tax reform. Alan Viard of the American Enterprise Institute focuses his analysis on the wealth tax proposals put forward by presidential candidates Senator Elizabeth Warren (D-Massachusetts) and Senator Bernie Sanders (I-Vermont). Viard argues that a wealth tax would likely reduce national saving and investment, although the latter impact would likely be ameliorated somewhat by an increase in capital inflows from abroad. He also highlights that although the proposed wealth tax rates appear to be low, they are equivalent to high-rate income taxes. Finally, he cautions that a wealth tax would potentially encounter a number of compliance and administration hurdles, and its revenues would likely erode if Congress were to add additional exemptions for certain asset classes, such as real estate.

Lily Batchelder, the Robert C. Kopple Family Professor of Law, and David Kamin, Professor of Law, both of NYU School of Law, analyze the benefits and drawbacks of four different but potentially complementary approaches to increasing the progressivity of the tax code and raising drastically more revenue from the wealthiest Americans. These approaches include a dramatic increase in the progressivity of taxes on labor and other ordinary income; a tax on accrued gains of the wealthy at ordinary income tax rates; a wealth tax on high-net-worth individuals; and a financial transactions tax. They also identify incremental approaches that could be taken to

---

raise more revenue from the wealthy without making substantial changes to existing tax structures, including the reversal of several provisions in the 2017 tax law, taxing accrued gains at death at higher rates, and reducing the exemption threshold and increasing the rate of estate taxes, among other approaches.

**Conclusion**

Government in a market-based, liberal democracy should seek to foster healthy market competition, promote economic opportunity, and address inequities in the market distribution of outcomes in ways that are consistent with our democratic ideals and values. This volume represents an effort to bring the best evidence to bear on some of the greatest economic challenges facing American capitalism today and on potential policy responses to those challenges.
PART I

MARKET CONCENTRATION

Causes, Consequences, and Policy Responses to Market Concentration
Thomas Philippon

Concerns About Concentration
Nancy L. Rose
Causes, Consequences, and Policy Responses to Market Concentration

AUTHOR

Thomas Philippon*

ABSTRACT

I review the causes and consequences of rising concentration of market shares that is occurring in most U.S. industries. While concentration is not necessarily harmful to the economy, my assessment of the available evidence leads me to conclude that rising market concentration since the early 2000s has produced market inefficiencies. Increased barriers to entry have resulted in lower investment, higher prices, and lower productivity growth. I estimate that the associated decline in competition has likely decreased aggregate labor income in the United States by more than $1 trillion between 2000 and 2019. Policy responses should include regulatory reform, which in some instances will mean less regulation (e.g., less occupational licensing that protects incumbents in an industry) and in other instances increased regulation (e.g., with regard to business practices of two-sided digital platforms) and a renewed focus on antitrust enforcement. Finally, I review specific actions that should be taken to promote competition among two-sided, internet platforms.

* New York University, the Center for Economic and Policy Research, and the National Bureau of Economic Research. tphilipp@stern.nyu.edu. This memo builds on a research project with German Gutierrez, Callum Jones, and Matias Covarrubias, and on an extensive analysis of the U.S. economy over the past 20 years from my forthcoming book, The Great Reversal. I have benefitted from feedback and discussions with many people, Janice Eberly and Chad Syverson in particular. I am grateful for financial support from the Smith Richardson Foundation.
1. Introduction

The American economy is less dynamic than it used to be. Fewer jobs are being created or destroyed, fewer new firms are starting, and fewer old firms are dying (Haltiwanger, Jarmin, & Javier, 2011). The secular decline in market dynamism across U.S. sectors was first observed in the retail sector during the 1980s and has since spread to nearly all sectors of the American economy, including the high-growth, information technology sector (Decker, Haltiwanger, Jarmin, & Miranda, 2015). Declining dynamism is a troubling sign for the health of the American economy, as it may signal that market leaders have gained the ability to maintain their market position at the expense of competition, innovation, and higher productivity growth. The decline in market dynamism also coincides with increasing profits and concentration across industries (Council of Economic Advisors, 2016; Grullon, Larkin, & Michaely, 2019) and a falling labor share of income (Elsby, Hobijn, & Sahin, 2013; Karabarbounis & Neiman, 2014). Taken together, these trends suggest that rent-seeking and anticompetitive practices may be creating or reinforcing a less dynamic economy.

An important question for policymakers is the extent to which growing market concentration is a symptom of increased rent seeking, or the result of more benign factors, such as changes in the distribution of productivity across firms or a shift toward intangible assets with strong scale effects. The purpose of this memo is to shed light on this important question by reviewing recent trends in market concentration and documenting how the characteristics of many U.S. industries diverge from counterparts in Europe, Asia, and other non-U.S. markets. Finally, I discuss the policy responses to increasing market concentration.

1.1 Market Concentration Is Not Necessarily Bad

Increasing market concentration does not necessarily imply that competition is weakening. Trends in concentration reflect a complex interplay of market forces and regulatory actions, so we must better understand why a change in concentration has taken place in order to assess whether that change promotes or inhibits competition.

An increase in concentration can reflect healthy competition in the marketplace. When producers in competitive markets drive down profit margins and drive out inefficient producers, measures of concentration in that market will increase. For example, a new technology could make it easier for customers to compare prices and judge the quality of competing goods and services within an industry. Lower search costs make it harder for inefficient producers to survive. Eventually, less efficient producers will exit the market or merge with another firm, which will cause measures of concentration in that market to increase. Measures of productivity will also increase because the remaining firms in the market are more productive, in
aggregate. The rise of superstores and e-commerce in the retail industry exemplifies this trend. Lower search costs weeded out inefficient retailers, which led to greater concentration, price competition, and aggregate productivity (Hortacsu & Syverson, 2015). A market with low search costs might therefore be a “winner take most” marketplace, as suggested by Autor, Dorn, Katz, Patterson, and Van Reenen (2017) and Van Reenen (2018).

Increasing differences between firms on measures of productivity and relative sales might also reflect capabilities—managerial, technological, reputational—that are not easily transferred between firms. These persistent differences might be embedded in intangible assets. Intangible capital might in turn explain the decline in private investment, as argued by Crouzet and Eberly (2018).

1.2 Rising Concentration May Signal Weakening Competition

An increase in market concentration can also signal weakening competition that results from lax antitrust enforcement and rising barriers to entry. In the absence of strong antitrust enforcement, rent-seeking behavior reduces investment and innovation both because the incentives to innovate are reduced and because the resources that firms expend to prevent other firms from entering or catching up to them could have been put to more productive uses.

Numerous studies demonstrate the association between recent increases in concentration and other indications of rising market power. Firms in concentrating industries have higher profits, positive abnormal stock returns, and more profitable mergers and acquisitions (Grullon et al., 2019). The decline in competition also plays a significant role in explaining the falling labor share of income (Barkai, 2017). De-Loecker, Eeckhout, and Unger (2018) argue that markups—which is the amount by which a firm can charge for a good or service over what it costs to produce the last unit of good or service—have risen substantially since the 1980s. Gutiérrez and Philippon (2017) show that rising concentration has led to weak investment. Gutiérrez and Philippon (2018) and Gutiérrez, Jones, and Philippon (2019) argue that domestic competition has declined in many U.S. industries because of increasing entry costs, lax antitrust enforcement, and lobbying. This has led to higher prices and lower investment. In an efficient economy, capital should flow to where it is most valuable. Gutiérrez and Philippon (2019) show that this has stopped happening in recent years: Rents are not competed away by free entry any more.

Nevertheless, there are cases in which it is less clear-cut whether an observed increase in market concentration reflects efficiency gains and enhanced productivity or more pernicious anticompetitive behaviors. In particular, the returns to scale associated with two-sided technology platforms, such as Facebook and Google, are difficult to characterize. On the one hand, industry leaders can become increasingly
more efficient than their followers. However, once an industry leader has gained a dominant position, they can exploit that position to prevent competitors from entering the market, and thereby increase their economic rents. Thus, the final assessment of the merits and drawbacks of concentration in the high-tech industry should depend on whether or not an increase in concentration is associated with rising wages, investment, and productivity growth.

2. Survey of Recent Trends

2.1 Aggregate Trends

Figure 1 illustrates that concentration and profits have increased in the United States, while the labor share of income and private investment have declined (Grullon et al., 2019; Autor et al., 2017; Gutiérrez & Philippon, 2017). Panel A shows the sales-weighted average change in the eight-firm Concentration Ratio (CR8), which measures the market share of the eight largest firms in an industry, across manufacturing and non-manufacturing sectors. Concentration has been growing at an accelerated pace in the non-manufacturing sector over time. Panel B shows that the profit rate, which is the ratio of after-tax corporate profits to value added (i.e., Gross Domestic Product), has risen from an average of 7% between 1970 and 2002 to an average of 10% in the period since 2002. Panel C shows the decline in the labor share of income over time by about 5 points of GDP. Finally, Panel D shows the ratio of net investment to net operating surplus. Firms used to reinvest about 30 cents of each dollar of profits. Now they only invest 20 cents on the dollar, which accounts for the weakness of corporate investment in recent years.

Figure 2 illustrates that these patterns are unique to the United States. Panel A shows that profits increased only in the United States while in Europe and other advanced Asian economies (Japan and South Korea) they remained stable or even declined. Panel B shows that concentration increased in the United States, while it remained roughly stable in Europe and Asia. Panel C shows that the labor share has declined in the United States since 2000, but it has remained stable in Europe. Assuming that all advanced economies use similar technologies, the uniqueness of U.S. trends suggests that technology alone cannot explain them.

1 For this figure, we measure concentration as the ratio of sales by the eight largest firms in Compustat that belong to a given KLEMS industry x region to total Gross Output reported in OECD STAN. Corporate consolidation is therefore accounted for, as dictated by accounting rules. The appendix provides additional details on the calculation, while Gutiérrez and Philippon (2018) provide a detailed comparison across a wide range of concentration measures for the United States and Europe. Bajgar, Berlingieri, Calligaris, Criscuolo, & Timmis (2019) use ORBIS data to include private firms; and take into account that some firms are part of larger business groups. When they measure concentration at the business group-level within two-digit industries, they find a moderate increase in concentration in Europe, with the unweighted average CR8 increasing from 21.5% to 25.1%. In North America, CR8 increases from 30.3%– 38.4%.
Figure 1. Evolution of U.S. Concentration, Profits, Labor Shares, and Investment

Notes: Panel A based on the cumulative sales-weighted average change in eight-firm Concentration Ratio (CR8). Data from the U.S. Economic Census based on SIC-four codes before 1992 and NAICS-six codes after 1997. We include only those industries that are consistently defined over each 5-year period, so that no change is measured from 1992 to 1997. When multiple tax groups are reported, only taxable firms are included. CR8 equals the market share (by sales) of the eight largest firms in each industry. Panels B, C, and D are based on quarterly data for the Non-Financial Corporate sector from the Financial Accounts of the United States, via FRED. Profit rate is defined as the ratio of After Tax Corporate Profits with IVA and CCAdj to Value Added (series W328RC1A027NBEA and NCBGVAA027S, respectively). Labor Share is defined as the ratio of compensation of employees (NCBCEPQ027S) to gross value added (NCBGVAQ027S). NI/OS is defined as the ratio of net investment (gross fixed capital formation minus consumption of fixed capital, series NCBGFCA027N minus NCBFCFCA027N) to net operating surplus (series NCBOSNQ027S). Dotted lines show the average of the corresponding series before and after 2002 (which is the year where we have the census concentration measure in Panel A).
2. Turnover Among Market Leaders Has Declined

Simple measures of market concentration cannot tell us whether concentration stems from good (more productive) or bad (less productive) factors. Instead, many economists prefer to use alternate measures of market power, such as the one proposed by Covarrubias, Gutiérrez, and Philippon (2019), which measures the turnover of market shares and market leadership within industries. In particular, it measures the probability that a firm that is at the top of its industry today—defined as being in the top four firms or top 10% of market value—will drop out of that position within the next 5 years. Figure 3 demonstrates that the likelihood of a leader being replaced was 35% in the 1980s, increased to 40% at the height of 1990s dot-com bubble, and subsequently declined to only 25% today.

Figure 2. Profits, Concentration, and Labor Shares Across Advanced Economies

Notes: Gross Operating Surplus divided by Productivity (GOS/PROD) and for Non-Agriculture business sector excluding Real Estate, from OECD STAN. Change in eight-firm Concentration Ratio (CR8) for Non-Agriculture business sector excluding RE, based on Compustat but adjusted for coverage using OECD STAN. CR8 for Japan + Korea reported only since 2006 because Compustat coverage increases rapidly beforehand. Change in labor share for Market Economy, from EU KLEMS. See data appendix for details.
2.3 Airlines and Telecoms: Weak Antitrust and Bad Concentration

Until the 1990s, U.S. markets were more competitive than European markets. Today, however, European markets have lower concentration, lower excess profits, and lower regulatory barriers to entry. Two U.S. industries in particular exemplify the evolution of concentration and markups over time: airlines and telecoms. Figure 4 plots the evolution of markups and concentration for the telecom and transportation/air industries, respectively. While the two measures exhibit little, even negative, correlation before 2000, both rise sharply after that year. This is consistent with the cross-country analyses of Gutiérrez and Philippon (2018), which show that concentration explains a significant share of price differences across countries in recent years.

The United States used to be a leader in internet access. Nearly 20 years ago, Economides (2002) observed that one of the key reasons for Europe’s lag in internet adoption was the fact that in most countries, unlike the United States, consumers were charged per minute for local calls. As a result, access to the internet was more expensive in Europe than in the United States.

Things have changed dramatically over the past 20 years, however. In 2018 the average monthly cost of fixed broadband was nearly twice as expensive in the
United States ($68) compared to Europe, where costs ranged from $30 to $40 in most countries (Best Broadband Deals, 2019).²

Air transportation is another industry in which the United States has fallen behind Europe in market competition. The rise in U.S. concentration and profits aligns closely with a controversial merger wave that included Delta-Northwest (2008), United-Continental (2010), Southwest-AirTran (2011), and American-US Airways (2014). Today, European airlines are far more competitive than U.S. counterparts along measures of both concentration and prices.

![Figure 4. Change in Markup and Concentration Since 1991: Airlines and Telecom](image)

**Source:** Compustat BLS multifactor tables for markups. Compustat for import-adjusted concentration.

### 2.4 Foreign Competition

Globalization has fundamentally reshaped the structure of American industries exposed to foreign competition. A large literature documents these effects. Capital-intensive plants and industries are more likely to survive and grow in the wake of import competition (Bernard, Jensen, & Schott, 2006). Chinese import competition leads to increased technical change within firms and a reallocation of employment toward more technologically advanced firms (Bloom, Draca, & Van Reenen, 2015).

² South Korea and Japan were similar to Europe. The authors of the report are puzzled by U.S. prices and conclude that “while broadband in the United States is widely available and uptake is high, lack of competition in the marketplace means Americans pay far more than they should, compared to much of the rest of the world.” Faccio and Zingales (2017) estimate that U.S. consumers would gain $65 billion a year if U.S. mobile service prices were in line with German ones.
Research and development (R&D)-intensive firms are better able to cope with Chinese competition than low-R&D firms (Hombert & Matray, 2014). Markups decreased in industries affected by foreign competition (Feenstra & Weinstein, 2017).

Common measures of concentration only include domestic firms, but this can be misleading when trade represents a significant share of an industry’s output. Figure 5 shows the normalized number of firms in industries with high- and low-Chinese import exposure relative to the year 1991. Both groups trend together before 2001, but start to diverge after that point, which is when China entered the World Trade Organization (WTO).

Global trade creates a challenge for naive measures of concentration. If we only count domestic firms, we could falsely conclude that concentration has increased when domestic concentration has simply responded to the increasing presence of foreign competitors in the United States. In Covarrubias et al. (2019), we present measures of concentration that control for imports. This trade adjustment plays a significant role in about half of manufacturing industries, which represent less than 10% of the economy. So, while adjusting for trade is important for those sectors, it does not change the conclusion that industries have become more concentrated.

Figure 5. Number of Firms by Chinese Exposure (1991=1)

Notes: Annual data. Number of firms from Compustat; import penetration (IE) based on NBER-CES and Peter Schott’s data. Manufacturing industries only, split into “high” (above-median) and “low” (below-median) exposure based on import penetration from 1991 to 2015. See data appendix for details.

---

3 We follow Autor, Dorn, Hanson, Pisano, and Shu (2016) and define import penetration for industry $j$ at time $t$ as

$$\Delta IP_{jt} = \Delta M^{UC}_{jt}$$

where $\Delta M^{UC}_{jt}$ denotes the change in U.S. imports from China from 1991 to $t$; and $Y_{jt} + M_{jt} - E_{jt}$ denotes the initial absorption (defined as output, $Y_{jt}$, plus imports, $M_{jt}$, minus exports, $E_{jt}$). $Y_{jt}$ is sourced from the NBER-CES database; while $M_{jt}$ and $E_{jt}$ are based on Peter Schott’s data. Only NAICS level six industries where data are available across all sources are included in the analyses. See also Pierce and Schott (2016).
2.5 Barriers to Entry Have Increased

The secular decline in firm start-ups has coincided with the sizeable increase in the number and stringency of federal regulations in the United States. Figure 6 uses data from the RegData database, which aims to measure regulatory stringency at the industry-level.\textsuperscript{4} It relies on machine learning and natural language processing techniques to count the number of restrictive words or phrases such as ‘shall’, ‘must,’ and ‘may not’ in each section of the Code of Federal Regulations and assigns each instance to an industry.\textsuperscript{5,6} Burdensome federal regulation can be problematic if market leaders are able to co-opt the regulatory regime to increase the barriers to entry in their industry.

\textbf{Figure 6. Regulation Index and Firm Birth Rate}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{Regulation Index and Firm Birth Rate}
\end{figure}

\textbf{Source:} Firm entry rates from Census’ Business Dynamics Statistics. Regulatory restrictions from RegData. See text for details.

\begin{itemize}
\item \textsuperscript{4} Introduced in Al-Ubaydli and McLaughlin (2015)
\item \textsuperscript{5} This represents a vast improvement over simple measures of “page counts,” but it is still far from a perfect measure. Goldschlag and Tabarrok (2018) provide a detailed discussion of the database and its limitations, including several validation analyses that, for example, compare RegData’s measure of regulatory stringency to the size of relevant regulatory agencies and the employment share of lawyers in each industry. Goldschlag and Tabarrok (2018) conclude that “the relative values of the regulatory stringency index capture well the differences in regulation over time, across industries, and across agencies.” One limitation is that the main RegData database covers only federal regulation. State and local governments also have regulatory responsibilities which further add to the regulatory burden. It is hard to summarize the scale or growth of state and local regulation, but the increase has also been significant. Occupational Licensing is an area that has received substantial attention.
\item \textsuperscript{6} One limitation is that the main RegData database covers only federal regulation. State and local governments also have regulatory responsibilities which further add to the regulatory burden. It is hard to summarize the scale or growth of state and local regulation, but the increase has also been significant. Occupational Licensing is an area that has received substantial attention. Council of Economic Advisors (2016), for example, show that the share of workers required to obtain a license increased from under 5% in the 1950s to over 25% in 2008—in large part because of greater prevalence of licensing requirements at the state-level.
\end{itemize}
2.6 The Growth of Intangible Assets?

Crouzet and Eberly (2018) argue that the rise of intangible capital—such as software, intellectual property, brand, and innovative business processes—can explain some of the weakness in physical capital investment since 2000. It is important to emphasize, however, that investment has been weak in all asset classes, including intangible assets. Figure 7 shows that the growth of the capital stock has fallen across all asset types since 2000. The shift toward intangible expenditures is clearly present across all advanced economies, as shown in Covarrubias et al. (2019). Profits, on the other hand, have only increased in the United States. In the other regions, they have remained flat or even declined.

**Figure 7. Growth Rates of Capital Stock**

![Graph of growth rates of capital stock](image)

**Notes:** Growth rate of private nonresidential fixed assets; based on section 4.2 of the BEA’s fixed assets tables
3. Policy Implications

3.1 From Good to Bad Concentration

Increasing concentration can be a good or a bad sign for the health of a competitive market. If it is good, then concentration should be associated with lower prices and higher productivity. If bad, concentration would be associated with higher prices and lower productivity. Covarrubias et al. (2019) construct a dozen indicators of prices, productivity, investment, and concentration and show that two main factors—called “principal components”—explain the evolution of these various indicators.

Figure 8 plots the scores of the two factors over time: The first measure (PC1) captures the data that is consistent with the idea that intangibles drive concentration. The measure will increase if intangible investment and productivity are higher. The second score (PC2) captures the data that is consistent with the argument that there are rising barriers to entry and weakening antitrust enforcement in the United States. We find that both theories are important for explaining the evolution of U.S. industries over the past 20 years, but the relative importance of each measure has changed over time. In the 1990s and until the early 2000s, we find that the intangibles explanation dominates. However, barriers to entry and weakening antitrust enforcement become increasingly important after the mid-1990s.
3.2 Welfare Losses From Lack of Competition

I estimate that markups in the United States have increased by about 12% since 2000 (see Appendix A for calculations). Such an increase in markups implies that wages and consumption are at least 10% below their potential. With a simple, back-of-the-envelope calculation, I estimate the amount by which higher markups have lowered labor income. Since U.S. GDP is about $20 trillion, and labor income is about 60% of GDP, labor income is about $1.44 trillion. Thus, increasing markups in the United States have lowered labor income by about $1.44 trillion. A return to the level of competition that prevailed in the United States in the late 1990s would add about $1.44 trillion to labor income in the United States.

3.3 Applying the Good Concentration/Bad Concentration Framework

The overarching goal of policy should be to let good concentration happen and to reverse or prevent bad concentration.

Policy should restore competition in markets where it has declined significantly (air travel, telecoms) and reverse the trend toward increasing barriers to entry and anticompetitive regulations at the federal and state levels.

The tools to do so are mostly regulatory, but antitrust policy could also play a role. The details vary from one industry to the next, but here are some concrete examples:

- **Airlines**: Successful reforms in recent years in Europe have improved the allocation of takeoff and landing slots. Slots are now reserved for newer, younger, and smaller carriers at major airports.

- **Regulations**: At the state level, there should be more legislation like the Ice Cream Freedom Bill. This refers to the bill passed in Arizona that loosens licensing requirements for mom-and-pop restaurants that make their own small-batch ice cream. Arizona recently changed its laws to recognize out-of-state occupational licenses for more than 40 professions, from cosmetologists to surgeons.

- **Telecoms**: French households, on average, can choose from five internet providers for their home internet. American households, on average, have 1.5 choices, that is, half of households have two choices, while the other half have only one provider available to them. See Gutiérrez and Philippon (2018) for a discussion of reforms in Europe.

- **Legal professions**: Licensing and other requirements in the U.S. legal profession prevent markets from being competitive. Bar requirements should be transferable across U.S. states.
• Health care: The health-care industry is rife with concentration that reflects anticompetitive practices that are driven by an arms’ race mentality for greater market power. Each player in the industry (drug makers, insurers, health providers) reacts to an increase in market concentration among industry counterparts by lobbying to obtain even greater market power for itself.

These examples are relatively straightforward in the sense that it is relatively easy to argue that concentration in these industries has led to losses in consumer welfare.

As we have seen, however, productivity gains and winner-take-all dynamics can lead to an efficient form of concentration. This is more likely to be the case when concentration comes together with large investments in intangible assets and technologies that feature increasing returns to scale. This is also why the analysis of digital platforms is complex and requires a separate discussion.

3.4 Concentration and Competition in the Digital Economy

Can we apply the good vs. bad concentration framework to the digital economy? Yes, but we need to clarify two important issues.

First, we must clarify whether a consumer welfare paradigm is still the right one to apply to two-sided, technology platforms such as Facebook, Amazon, and Google. These firms are described as “two-sided” because they do not charge consumers for the use of the platform, but rather earn revenue by charging merchants (e.g., advertisers, retailers, etc.) for access to consumers via their platform. A naive interpretation of the consumer welfare standard would lead one to conclude that consumers cannot possibly be worse off since access to the platform for consumers is free, many services are also offered to consumers for free (e.g., Gmail), and the price of goods that are sold are often lower than those of competitors (e.g., Amazon Prime).

However, it would be incorrect to equate consumer welfare with short-term price decreases. In standard economics, consumer welfare is the sum of all future consumer surpluses. If a firm engages in predatory pricing today to later raise prices, consumer welfare is likely to decrease over the long-run. When a platform uses its monopsony power against producers, it limits their incentives to invest and innovate, and thus consumer welfare declines. For instance, Amazon monitors its own marketplace to determine which products are popular; it has the opportunity to use this data to introduce copycat products that undercut the original seller. This could deter producers from innovating, and in that case, it could result in lower consumer welfare. Policymakers should seek to maximize consumer welfare, as defined in the economic sense rather than in a naive legal sense that only considers short-run price changes.

Second, the stars of the digital economy—Amazon, Google, Facebook, Apple, and Microsoft (GAFAMs for short)—are not as “special” as one might think. In my book
Part I: Market Concentration

(Philippon, 2019), I compare the star firms of the U.S. economy from each decade since World War II. I find that there have always been stars, and they have always been productive, innovative, and profitable. Along all quantitative dimensions, including profit margins and productivity, the stars of today look quite similar to the stars of the past. If anything, they are smaller than market leaders of the past, and they matter less for overall GDP growth than General Motors, IBM, or AT&T did at their peak. An important implication of these facts is that, to paraphrase Jane Austen, the GAFAMs deserve neither such praise nor such censure. As far as regulations and antitrust policy are concerned, they should be assessed and treated just like other firms.

3.5 Good Concentration/Bad Concentration in the Digital Economy

Concentration in the digital economy creates two conceptually separate issues: One is market dominance; the other is privacy. The stars of the digital economy—Amazon, Google, Facebook, Apple, and Microsoft—combine features of good and bad concentration as defined above, which contributes to the controversies surrounding these firms. On the one hand, they clearly enjoy high market shares and outsized profits and work aggressively to maintain their dominant positions. The platforms they operate are less than transparent and present the potential for conflicts of interest when the firm running the platform also sells goods and services on that platform. On the other hand, they are very innovative and they “give away” some services for “free,” adding to short-term consumer utility as discussed above.

The range of options for addressing increased market dominance of these firms is quite broad, ranging from a laissez-faire approach to breaking up these firms. A laissez-faire approach might be in the best interest of U.S. policymakers, since these firms are profitable for U.S. shareholders but charge high markups in foreign markets, which results in a welfare loss for foreign consumers. Of course, foreign governments are likely to object to the United States taking a laissez-faire approach, rational as such an approach might have been until recently.

At the other extreme, some advocates argue technology monopolies should be broken up. This approach presents two challenges: First, the goal of such a break-up needs further clarification. Is the goal to address market dominance or privacy concerns? Second, breaking up these firms would take a long time.

There are several options that fall on the spectrum between laissez-faire and break-ups that policy makers should consider. These options are similar to those taken in the telecom industry in earlier decades.

- First, platforms should be required to be interoperable, which refers to the ability of a platform to exchange information with other networks. For example, a user cannot easily move information between their Facebook and Twitter profiles. This is analogous to the requirement that customers of
one phone company be able to place a call to customers of another phone company. Today, we take for granted that any phone user can call any other phone user anywhere, but this outcome did not happen spontaneously. It was mandated by regulators. The lack of interoperability between networks today is often a choice made by the dominant network to protect its dominant position.

- Second, data portability should be expanded. Users should be able to move their data from one network to another network. This is the equivalent of the rule that phone users may switch between providers while keeping their phone numbers, which lowers switching costs and increases competition.

- Third, users should have the ability to opt out of horizontal tracking. Google and Facebook are the only companies able to track users on millions of websites today, and this gives them a rich trove of data about each user that can be used to improve ad targeting, giving platforms a significant advantage over other online sellers. Users should be allowed to opt out of horizontal tracking on third-party websites.

- Fourth, conflicts of interests should be minimized across platforms, as Lina Kahn recently called for (2019). In the case of Apple, the controversy concerns its App Store, where prices are hidden and rules are obscure. In the case of Amazon, the platform serves simultaneously as a market place and a merchant, which makes it easy for the company to undercut the prices of competitors or give more favorable product placement to its own products. In the case of Google and Facebook the debate concerns the platforms’ power to prioritize some websites over others.

The final point to keep in mind is that competition matters for privacy because of the incentives that it creates. It is not by chance that Facebook started to disregard privacy issues when it felt assured of its monopoly power. Facebook initially promised not to use cookies to track people on third-party sites but reneged on that promise when it became the only dominant social network. The same applies to the other large digital firms. Competition implies that customers can vote with their feet—if they are dissatisfied with privacy issues on one platform they can switch to another. This threat provides a powerful incentive for firms to protect the privacy of their users.

4. Conclusion

Simple measures of industry concentration reveal that the market shares of dominant firms are increasing in most U.S. industries. While concentration is not necessarily harmful to the economy, my assessment of the available evidence leads me to conclude that rising market concentration since the early 2000s has produced market inefficiencies. Dominant firms have succeeded in increasing barriers to
entry, which has resulted in lower investment, higher prices, and slower productivity growth. Policy options should include deregulation at the federal and state level and a renewed focus on antitrust enforcement. In the case of two-sided, internet platforms, the starting point should be enforcing regulations that would give users greater control of their data and promoting the interoperability of platforms, which would be analogous to regulations in the telecom industry.
Appendix

To estimate the size of the markup increase over time, consider a standard profit-maximizing economy, and rewrite the markup $\mu$ of price $P$ over marginal cost $MC$ by multiplying and dividing by average costs:

$$
\mu = \frac{P}{MC} = \frac{P}{AC} \cdot \frac{AC}{MC} = \frac{AC}{MC} \cdot \frac{Revenue}{Cost}
$$

The ratio of average to marginal costs, $\frac{AC}{MC}$, equals the returns to scale for a cost-maximizing firm taking factor prices as given while $\frac{Revenue}{Cost}$ can be written as $\frac{1}{1-s_\pi}$ using the profit share in revenues $s_\pi$. There are many estimates of these numbers in the literature. In Covarrubias et al. (2019) we argue that returns to scale have increased by about 5% (say from 1 to 1.05). On the other hand, profits shares of revenues have increased from 2% to 8%. Thus:

$$
\frac{\mu_2}{\mu_1} = \frac{1 - 0.02}{1 - 0.08} \cdot 1.05 = 1.1185
$$

The data is thus consistent with an increase in markups by about 12%. This is also in line with a direct comparison of prices and unit labor costs between Europe and the United States, as documented in Gutiérrez and Philippon (2018).
References


Best Broadband Deals June 2019 | Compare Broadband Offers - Cable.co.uk. (n.d.). Retrieved June 28, 2019, from Cable website: https://www.cable.co.uk/broadband/


Concerns About Concentration

AUTHOR
Nancy L. Rose*

ABSTRACT
Recent reports that document increases in aggregate concentration measures and the correlation of concentration with a variety of other economic outcomes—including rising profit rates and a declining labor share—have led to great concern about the health of competition in the United States and global economies. This memo attempts first to clarify evidence on concentration trends in product and labor markets and to highlight significant measurement and interpretation challenges for aggregate studies of concentration. I then review the state of U.S. competition policy, focusing on impediments to rigorous enforcement of both merger policy and limits on exclusionary behavior. Finally, I suggest potential policy reforms that would promote competition, and describe some of the promises and pitfalls of these approaches.

* Massachusetts Institute of Technology: nrose@mit.edu
1. Introduction

There has been an explosion of concern in recent years about the state of competition in the United States and global economies. Media reports have highlighted growing concentration across industries and the dominance of large digital platforms; government agencies and non-governmental organizations have weighed in on the state of competition; economic researchers have documented trends in aggregate concentration measures and the correlation of concentration with a variety of other economic outcomes, including rising profit rates and declining labor share of income; and some politicians have taken up calls for policy changes to invigorate antitrust enforcement, regulate dominant firms, or even break up large tech companies (Graham, 2019). These debates extend to labor markets as well as to product markets, with an increasing body of research exploring labor market power, or so-called “monopsony.”

In this memo I first summarize recent evidence on concentration trends in product and labor markets, highlighting significant measurement and interpretation challenges for aggregate studies. I then briefly discuss U.S. competition policy, focusing on challenges to rigorous enforcement of both merger policy and what antitrust practitioners call “unilateral conduct” (such as monopolization or exclusionary behavior). Finally, I suggest potential policy reforms to preserve or increase competition, describing some of their promises and pitfalls.

2. Industry Concentration: What Can We Make of Reported Trends?

2.1 Issues in the Measurement of Industry Concentration

There are dozens of recent studies attempting to measure economy-wide changes in industry concentration over the past several decades. Many of these studies then relate the measured changes in concentration to outcomes such as corporate profits, markups, or labor share. The authors of these studies use a variety of different data sets and methodologies to measure concentration, some more convincing than others. I

1 For example, “Too Much of a Good Thing” (2016); “The Next Capitalist Revolution” (2018); Porter (2016); Francis and Knutson (2015); Ip (2019).

2 For example, Council of Economic Advisors (2016a, 2016b); “1st Joint IMF-OECD-World” (2018); Bajgar, Berlingieri, Calligaris, Criscuolo, and Timmis (2019); “The Rise of Corporate” (2019).


4 See Azar, Marinescu, and Steinbaum (2017); Azar, Marinescu, Steinbaum, and Taska (2018); Benmelech, Bergman, and Kim (2019); Prager and Schmitt (2019); and Rinz (2018).
begin this discussion by laying out some principles to guide judgement about which measures of industry concentration are likely to yield the most meaningful statistics:

(i) *Industries should be defined narrowly.*

Economy-wide concentration studies typically use the North American Industrial Classification System (NAICS) or Standard Industrial Classification (SIC) codes to define industries, with levels of aggregation that range from very broad one- or two-digit sectors (“Manufacturing”) to more narrow four- (SIC) or six-digit (NAICS) industry-specific codes (“Breakfast Cereal Manufacturing”). Aggregations less specific than the four-digit SIC industry code are almost surely too expansive to provide insight into anything beyond the question of whether large firms in broad sectors are getting larger. As an example, the NAICS three-digit “Food Manufacturing” industry comprises manufacturers of breakfast cereal, chocolate and confectionary, dog and cat food, and animal slaughterhouses, among many, many others. It is difficult to think of what one could learn from changes in firm revenue shares, let alone concentration, across this mix of activities. The specificity of four-digit SIC or six-digit NAICS codes generally produces more interpretable industry definitions, though even these are rarely well-defined markets from a competitive standpoint.

(ii) *Measures of revenue shares should be built up from establishment data, not from assignment of top-line, firm-level sales.*

Some studies in this literature rely on firm-level databases, such as Compustat, that report a primary industry code for a firm, typically at a four-digit SIC level. The assignment of all of a firm’s revenue to one code in most cases systematically biases measures of industry concentration upward. It is much more accurate to measure industry revenues in the United States using the establishment level data produced by the Economic Census.

---

5 See Appendix A1 for examples. Some data providers append additional digits to the SIC or NAICS codes to create finer product level distinctions. While not part of the official classification system, this may allow finer gradations in classification.

6 While overly broad levels of aggregation likely on average underestimate concentration in true markets, they are uninformative with respect to concentration levels in any particular market. Moreover, changes in concentration may reflect compositional changes in industry mix without any change in the concentration of any individual market.

7 This is because many firms, particularly the largest ones, operate across several different industry segments. For instance, Autor et al. (2019) report that in 2012, the largest firm in a given four-digit industry operates in an average of nine other four-digit industries (down from 13 in 1982), and one-quarter of top four firms in one industry are among the largest four firms in another four-digit industry.

8 Some studies have used establishment-level data available from NETS, a privately produced data set; see e.g., Rossi-Hansberg et al. (2019). See Bajgar et al. (2019) for a discussion of the myriad problems with Orbis, used for many studies of European Union concentration.
(iii) Concentration measures should be based on the universe of firms, not only large or publicly traded firms.

Economic activity in many small and privately held firms may be missing in databases such as Orbis or Compustat, which rely on publicly reported financial data, such as 10-Ks, that privately held firms may not disclose. If the total commerce in these firms is significant, individually or in the aggregate, statistics excluding their activities may be misleading and will distort changes when the companies sampled change over time. This is especially problematic for studies of European concentration based on Orbis data, which expanded coverage of small and midsize European companies over time.

(iv) Concentration measures should reflect the size distribution of firms.

Industrial organization economists and antitrust practitioners prefer the Hirschman-Herfindahl Index, or HHI, which is the sum of squared market shares of all firms. Higher HHIs reflect more concentrated revenue, with an upper limit of 1 (or 10,000, if shares are measured as 0-100%) for monopoly. This provides more information about revenue distribution than concentration ratios, which are the revenue share accounted for by the largest N firms (commonly four or eight, denoted as CR4 or CR8). For example, a CR8 of 80% could reflect one firm with a 75% share, or eight firms with 10% shares, with very different implications for market structure. The HHI would distinguish between these situations.

(v) Concentration measures should reflect the appropriate geographic scope of a given product market.

This is aspirational and is virtually never satisfied in aggregate studies of concentration. Almost all studies apply a single geographic aggregation, typically national, to all industries. This is too narrow for markets with globally traded goods, such as aircraft, cement, or petroleum, and much too broad for markets with locally delivered goods and services, such as scheduled airline service between cities, concrete, or retail gasoline. Furthermore, to the extent that imports or exports are important in a given market, measures built up from sales only by U.S. entities could have severe mismeasurement. This is also problematic for firm-level data sources like Compustat, for which U.S. sales may be a fraction of firms’ recorded global revenue.

2.2 Four Main Takeaways From the Literature on Market Concentration

I offer four main takeaways from the burgeoning literature on trends in market concentration. My critical read of the literature incorporates the measurement principles described above to interpret and prioritize various studies.

(i) Studies of broad industry categories at the national level suggest increased concentration of revenue among the largest firms over the past 20 to 40 years.

The work of Autor et al. (2019) is representative of estimated trends in average
concentration levels built from establishment-level data for four-digit SIC industries at the national level. Figure 1 reproduces the figure from Autor et al.\(^9\) that graphs CR4 and CR20 for revenue and employment concentration.\(^{10}\) While all sectors show average increases in the CR4 between 1982 and 2012—between 5- and 15-point increases in the CR4—the rates of increase vary considerably. The smallest increases are in manufacturing, for which many product markets are more likely to be national or global in scope. The average manufacturing industry evidences a 4-point rise in the CR4, to just under 44%, which would be consistent with an increase in average firm share from 10% to 11% for each top four firm over the 30-year period. This is about the same increase as in Services, where the level of CR4 is much lower, reaching less than 15% in 2012. Retail trade experiences the greatest increase, roughly doubling the CR4 over 20 years, from 15% to 30%, for an average share of 7.5% for each of the top four firms. Finance, Utilities and Transportation, and Wholesale Trade experience increases between these endpoints, but only Utilities and Transportation end up with four-firm levels of concentration as high as 40%. If the SIC4 industries in this figure were true markets, it would not seem that concentration at any of these reported levels should trigger alarm, as the CR4 statistics suggest no fewer than 9 (Manufacturing) to 26 (Services) competitors in the average individual industry. On the other hand, too broad a definition could mask significantly higher concentration in more narrow product or geographic markets.

Figure 2 reproduces the Autor et al. (2019) graphs on concentration as measured by the HHI. This is scaled between 1 for an industry with 100 firms, each with a revenue share of 1%, to 100 for a monopoly. In none of the sectors is the average industry even moderately concentrated. The Horizontal Merger Guidelines (HMGs) used by the United States Department of Justice Antitrust Division and the Federal Trade Commission to evaluate mergers specify markets with HHIs greater than 18 on this scale to be moderately concentrated; those with HHI above 25 on this scale are highly concentrated.\(^{11}\) The highest average HHI in Figure 2 is between 8 and 9, in Manufacturing and Utilities and Transportation, which would be the value for an industry comprised of 11 to 12.5 equal-sized firms. In Services, the average HHI doesn’t even reach 2. A second difference from Figure 1 is that in both Manufacturing and Wholesale Trade, the revenue-based HHI is virtually the same in 2012 as it is in 1982, in contrast to the CR4 results. At this level of aggregation, even the broad conclusion that concentration has increased is sensitive to seemingly innocuous measurement choices.

---

\(^9\) Figures 1 through 3 are reproduced from academic papers. I apologize that these are not formatted for readability in black-and-white reproduction, red/green colorblindness, or other limits discerning color-based distinctions.

\(^{10}\) These are constructed from establishment-level data from the U.S. Economic Census, aggregating revenue to the firm-industry-year to compute concentration ratios and HHIs (scaled 0 to 100) at the four-digit SIC level. Industries are weighted by employment to compute the six broad sector averages graphed in these figures. The axes are not standardized across sectors, so neither the level nor the slopes of the curves (reflecting the rate of increase over time) are comparable across the sectors.

\(^{11}\) In antitrust, HHIs are measured with shares between 0 and 100, so the HHI ranges from 0 to 10,000 for a monopoly. The 2010 HMGs define cutoffs of 1800 for moderate and 2500 for high concentration.
Figure 1: Average Concentration (CR4, CR20) in Four-Digit Industries by Sector.

Source: Autor et al. (2019), Figure 4.

Note: Top (blue and green Lines with circles) are revenue shares; bottom (red and orange lines with triangles) are employment shares. Top four firm shares are plotted on the left axis, top 20 on the right.
Figure 2: Average U.S. Concentration (HHI) in Four-Digit SIC Industries by Sector, 1982-2012.

Source: Autor et al. (2019), Appendix Table A.1.

Note: The Hirschman-Herfindahl Index (HHI) is scaled 0 to 100. The blue circles plot the HHI calculated using firm sales and the red triangles plot the HHI calculated using employment.
While these particular figures focus on U.S. markets, there is evidence suggesting these trends are shared in other developed economies, perhaps with somewhat higher increases in CR4-type measures in the United States. It is difficult to access comprehensive microdata outside the United States, which can affect cross-country comparisons. While the similarity of results continues to be debated, it is likely a mistake to think that whatever explanation accounts for these trends should have a U.S.-centric focus.

(ii) Rising national concentration is not mirrored by increased concentration at the more local level, which recent work suggests has declined on average. A plausible explanation for this divergence is growth in the national revenue share of the largest firms in most industry categories, accompanied by expansion of those firms into new geographies.

Defining industry boundaries is only part of the challenge of defining a market in which firms compete. Geography also plays a critical role. Consider two industries in the NAICS segment 3273. Cement (NAICS 327310) is manufactured centrally and transported long distances, even internationally, particularly where low-cost water transport is available. Concrete (NAICS 327320)—a mixture of cement, aggregate (gravel or sand), and water—must be consumed within about 45 minutes of mixing, sharply limiting the market radius of a ready-mix concrete plant. This distinction is important. A U.S.-wide market for cement that excludes imports from industry sales is likely too narrow and may make the industry look more concentrated than it actually is if the imports are produced by non-U.S. firms and if U.S. exports are low. But, aggregating concrete revenue to the national level likely makes the concrete industry look much less concentrated than are the true local markets. Moreover, mis-aggregation can turn the implication of changes in concentration upside down. For example, a merger of a concrete firm operating only in the Southwest with a firm operating only in New England would appear to increase U.S. concrete industry concentration, even if, in the aftermath of the merger, local market concentration was unchanged, or perhaps lower if the firm opened up new concrete production facilities in previously unserved local markets.

Research by Rossi-Hansberg et al. (2019) and Hsieh and Rossi-Hansberg (2019) illustrates the practical importance of considering geography when trying to

---

12 Compare, for example, Autor et al. (2019) and Baiggar et al. (2019) to Gutiérrez and Philippon (2018) and Covarrubias et al. (2019). The disagreement over United States vs. European trends may be in substantial part dependent on the data sources used. See Baiggar et al. (2019) for a discussion of the impact of changes in Orbis coverage of small and mid-size EU firms over time and errors in firm ownership. Based on their corrections to Orbis, Baiggar et al. (2019) report increasing concentration (measured by CR8, the revenue share of the eight largest firms) in both the United States and Europe, although the magnitude of the increase is somewhat higher in the United States.

13 In the United States, imports account for roughly 10% of total consumption; see Portland Cement Association (2016).
understand the nature of economy-wide changes. Their work shows that national and local concentration trends between 1990 and 2014 diverge across most sectors of the U.S. economy. This is apparent in Figure 3, which compares average changes in narrow industry HHIs (scaled 0–1.0) at the national level to those for the same industries at the local level of eight-digit ZIP codes; similar results are obtained for counties or CBSA metropolitan areas. The results replicate the qualitative findings of increasing concentration at the national level over time, but local market trends are negative—in some industries, like Retail Trade, FIRE (Financial, Insurance, and Real Estate), and Services, very substantially so.

**Figure 3. Diverging National and Local Concentration Trends, Averaged to Sector Level**

Source: Rossi-Hansberg et al. (2019), figure 4.

Note: Average change in revenue-based HHI (scaled 0–1.0), computed from NETS establishment-level data. Industries are defined at eight-digit SIC level, and changes in HHI for each industry-geography are averaged with weights given by employment share of industry-location pair. Alternative geographic aggregations include: National, Core-Based Statistical Area (similar to MSA), County, and eight-digit ZIP code. Excludes industry-location pairs with no observations. Patterns are qualitatively similar using a balanced sample of industry-geography pairs observed for all years, although local concentration is roughly unchanged for manufacturing and wholesale trade by the end of the sample period.

14 These are based on establishment-level data from a private data source, National Establishment Time Series (NETS), which enables the researchers to observe sales in each year at fine levels of industry and geographic disaggregation. The industry definition is SIC8 (four-digit SIC with a product code appended). The authors exclude inherently “location specific” activities, such as agriculture, mining, or public utilities, where establishments may be constrained in location by natural resources or proximity to customers. Some of these (utilities) are likely local markets, in others (agriculture, mining) firms compete globally. See Rossi-Hansberg et al. (2019) for detailed discussion.
Concerns About Concentration

The authors offer a reconciliation for these seemingly contradictory trends. They present additional evidence that, within industries, the largest firm grew over time, in terms of both industry revenue share and locations served. It appears that the growth of the largest firms, on average, contributes both to rising national concentration and falling local concentration. By entering into new markets, these large firms bring an additional firm to a local market, thereby reducing local-level concentration. The authors confirm that when a large firm opens an establishment in a new ZIP code, local concentration falls and remains low over time.

In some industries, the rising national concentration is most relevant and the geographic dispersion of establishments that all compete in a regional or national market may offer little or no additional benefit to consumers. In other industries, particularly in the service sector, declining local concentration likely indicates more choices for consumers.

(iii) Aggregate estimates of average markups or profit rates appear to have increased over time.

Like the concentration literature, a large number of papers have tackled the question of whether aggregate profit rates, or markups of prices over marginal costs, have increased, and if so, by how much. Most, but not all, of the work reports rising markups, often of incredibly large magnitudes. For example, De Loecker, Eeckhout, and Unger (2018) report a tripling of the average margin in the United States from 20% over cost in 1980 to 60% over cost in 2016; De Loecker and Eeckhout (2018, p. 6) report that the “evolution of markups is comparable in Europe, North America, Asia and Oceania,” with increases of 40 to 60 percentage points. Autor et al. (2019) find that some production function-based estimation methods suggest markups increasing from 150% to 300% over this period.

The distribution of markups also has changed. Autor et al. (2019) report that when average markups are measured as the median markup or as unweighted average markups, only modest increases are observed. Substantial increases are observed in the average markup when it is weighted by firm value-added. This indicates either rising market shares of high-markup firms, growing markups for larger firms, or both. Autor et al. conclude that the higher than average markups over costs for the largest firms reflects their greater productivity relative to other firms in their industry category. They label these “superstar” firms, for their combination of scale, inferred efficiency, and margin levels.

A key question for scholars and policy makers is whether the calculations showing increased markups are reflective of increased economic rents, as many are inclined to assume. The implausible magnitude of many estimates, considerable sensitivity of implied markups to alternative estimation methods, and identified difficulties with some of the methods used to generate these numbers suggest some circumspection.
Is it credible that weighted average economic margins have increased from 20% to 60% between 1980 and 2016, as in De Loecker, Eeckhout, and Unger (2018)? Or from 120% (more than double marginal costs) to 200% (more than triple marginal costs), as some estimates in Autor et al. (2019) suggest? Those are astonishing numbers that yield implications inconsistent with other data on the economy (Basu, 2019).

Markups derived from accounting data are susceptible to a broad range of difficulties in mapping accounting data to economic costs and profitability. Capital cost accounting can be notoriously unhinged from economic costs, as discussed by industrial organization scholars of the 1970s and 1980s (e.g., Fisher & McGowan, 1983). These problems are compounded for companies that have significant investment in intangible capital such as intellectual property, information technology, advertising, research and development, and the like. Bessen’s (2017) research suggests that proprietary IT investment generates competitive advantages that give rise to both increased concentration and increased productivity, yielding higher estimated markups that could reflect normal returns to IT investments. These make accurate estimation of economic margins difficult and interpretation of estimates fraught.  

Basu (2019) describes the strength and weaknesses of the various approaches taken in the literature and highlights inconsistencies between the implications of estimated markups and observed patterns in macroeconomic data. He is reluctant to endorse any of these estimates, concluding that more research is needed to understand “why most markup estimates based on micro data are implausibly large and grow too fast in relation to the macro facts to be explained” (Basu, 2019, p. 20).

(iv) There is vigorous debate over the implications of these patterns in aggregate concentration and markups for the state of competition.

There are reasons to be cautious about concluding that market concentration has risen or is a meaningful problem for market competition and consumer welfare. Few of the existing studies that find increased market concentration calculate concentration at the level of a recognizable market. Markups and profit rates are difficult to measure with reliability and even more challenging to interpret. Furthermore, a long-standing literature casts significant doubt on the idea that cross-industry correlations of concentration with various outcomes imply reduced market competition.  

As that literature emphasizes, concentration is not necessarily the inverse of competition, and measuring the correlation of concentration and markups does not aid with the diagnosis. This is because changes in concentration measures have

15 Approaches that infer markups from production function-based estimates generally use stylized functional forms estimated at highly aggregated levels (for example, two-digit SIC industries). Even firms in the same narrow market exhibit substantial heterogeneity in productivity (e.g., Syverson, 2018), so imposing a common production function across two-digit sectors is more than heroic. Moreover, production function-based estimates also are sensitive to how cost data are reported and used, particularly assumptions about how reported accounting costs map into variable and fixed cost components.

no direct relationship to changes in market competitiveness or performance. For example, a market may become more highly concentrated when a firm acquires a competitor or increases barriers to entry, reducing competition and raising equilibrium prices. A wealth of detailed studies demonstrate these anticompetitive effects in the context of mergers across a broad variety of markets.\textsuperscript{17} If a merger reduces competition in input markets, such as labor, the firm may exercise its new market power by depressing what it pays workers or other suppliers (Prager & Schmitt, 2019), raising \textit{measured} markups in product markets and creating competitive harm upstream. Firms may soften competition and increase equilibrium prices by requiring trading partners to sign most favored nations clauses to ensure rivals cannot undercut them or by adopting customer loyalty programs such as frequent flyer rewards that make consumers unwilling to switch firms for modestly lower prices. In cases like these, higher concentration and adverse consumer impacts are outcomes of reduced competition.

In contrast to the above examples of anticompetitive behavior, a market might instead become more highly concentrated when one of the firms in that market becomes more efficient, enabling it to reduce prices and increase its market share, or when a firm develops an innovative product that consumers value, leading consumers to shift their purchases to that firm, perhaps even at a higher price, reflecting the greater consumer value (Demsetz, 1973). These cases may be associated with new capital, information technology, intellectual property, or other investments that reduce marginal costs or improve product offerings. Economists would characterize these markets as \textit{more} competitive, even though the outcome is associated with increased concentration and quite possibly both higher average markups and higher price associated with improved quality.

As another example, if fixed costs increase—for example, due to investments in information technology needed to produce a competitive product or consumer preferences for superstores with greater variety—average costs may increase and the equilibrium number of firms in a market may decline. This may generate a correlation between higher concentration, higher markups over marginal costs, and, depending on the context, even higher prices—but often also consumer benefits. These can all be outcomes of the competitive process, not a failure of it. Ganapati (2018) provides evidence of this phenomenon in wholesale trade, which has become much more concentrated in recent years as investments in information technology, logistics, and international supply and domestic distribution networks have facilitated the growth of the largest wholesalers. These wholesalers deliver greater variety and service to customers, reduce customer acquisition costs, and at the same time, realize higher markups from their “superstar” performance.

\textsuperscript{17} See Ashenfelter et al. (2014) and Kwoka (2015) on a broad set of merger retrospectives. Gowrisankaran, Nevo, and Town (2015), Ho and Lee (2017), Prager and Schmitt (2019), and the work reviewed in Gaynor, Ho, and Town (2015) analyze merger impacts on health care markets.
Perhaps surprisingly, there can be cases where a reduction in competition leads to a reduction in concentration, as can happen when a small number of dominant firms in a market tacitly or explicitly collude to raise their prices, ceding some of their collective market share to a group of fringe competitors while raising the dominant firms’ profits and reducing measured concentration. Miller and Weinberg (2017) show that in the aftermath of the Miller/Coors joint venture (JV), tacit collusion between Anheuser-Busch InBev and Miller/Coors increased. This led to rising prices and markups for their beer at the cost of eroding their market share in the years following the JV, reducing measured concentration over time.18

Finally, there may be markets in which firms compete to become large through innovative offerings that attract most consumers, generating competitive benefits. But if those markets then “tip” to insulate the market leader from any future competitive challenge, that same concentration may be associated with reduced competition and erosion of consumer value over time.

As these examples and a rich literature in industrial organization make clear, prices, profits, markups, and concentration are all codetermined outcomes of the competitive process in a market. There is not an independent causal relationship between concentration and prices or markups that can be inferred. Correlations measured across broad industries are particularly problematic, as there may be mixtures of each example above represented in the data.

As the examples mentioned here suggest, we can learn much more about competitive effects from detailed studies of individual industries that tackle issues of heterogeneity, causality, and competitive mechanisms head on. Deciding whether a policy intervention is needed, and if so, what it should be, requires solving those inference problems. Much of the recent literature focuses on economy-wide trends, and thus cannot deliver an accurate diagnosis of the issue. To take a medical analogy, a doctor’s decision to treat a fever with Tylenol, advanced antibiotics, or an emergency appendectomy depends on her diagnosis of the root cause of the fever. In any given situation, two of the treatments might prove both ineffective and costly to the patient. Similarly, economic policy prescription should be focused on treating the underlying causes, not simply symptoms.

3. Concentration in the Labor Market: What Should We Make of Reported Correlations With Workers’ Wages?

The literature on industry concentration trends developed in large part from an effort to understand the declining labor share of national income. A number of scholars have begun to focus directly on labor market concentration and outcomes for

---

18 The rise of consumer preferences for craft beer likely exacerbated the merger-induced decline in concentration.
Concerns About Concentration

This research correlates aggregate measures of employer concentration with wages, analogous to the literature correlating industry concentration with markups and other outcomes. Much of it concludes that occupations or industries in areas with fewer (more concentrated) employers are associated with lower wage levels. One should be cautious in assigning a causal relationship based on these studies, however.20

First, as with the measurement of market concentration in the product market, measurement of concentration in the labor market is fraught with issues and ambiguity. One issue is that when scholars attempt to define a “labor market,” they often define boundaries that do not align with relevant markets for employers or prospective workers. Studies that attempt to define labor market concentration are based on a variety of heuristics to facilitate regression analysis across many sectors and geographies. In some work (e.g., Benmelech et al., 2019), the labor market is defined as all workers in a particular industry-county pair, implicitly viewing workers in different occupations within an industry—manager, financial analyst, production line worker, custodian—as competing for jobs, but not companies in different SIC4 industries as competing with one another to hire mechanics or office managers. In other work (e.g., Azar et al., 2017; Azar et al., 2019), markets are defined by workers in the same six-digit Standard Occupational Code and commuting zone. This, for example, assigns four categories of “secretaries and administrative assistants”—Executive, Legal, Medical, and All Other—to four non-competing labor markets. Concentration measures in a number of studies are derived from vacancy postings by firms in a given occupation-commuting zone in a given quarter, a potentially narrow and noisy measure of employers.

Second, though most studies report a negative correlation between measures of labor market concentration and workers’ wages, they shed little light on the underlying reasons why wages are inversely related to employer concentration, even if one takes those correlations at face value. Without knowing why these two factors are negatively correlated, it is not clear what the policy implications are. There could be several alternative and inconsistent explanations, many analogous to concerns raised about the statistical studies of price (or markups or profits) and concentration in product markets described above.

Consider one example of variation driving changes in wages and changes in employer concentration in local geographies. Suppose a U.S. industry—say, automotive parts—shrinks or moves offshore, perhaps due to import competition. When one of these plants shut down, there is likely to be less overall demand for labor in its local market. Wages in that local market will likely fall, whether the labor

19 A sample of recent academic papers on the correlation of wages and employer concentration include: Azar et al. (2017), Azar et al. (2018), Benmelech et al. (2019), Rinz (2019).

20 This section draws heavily from Rose (2019).
market is perfectly competitive or not. The closure also is likely to increase the concentration of employment among the remaining employers, creating an inverse correlation between wages and concentration. But that is only a correlation; the root cause is the demand shock. Similar spurious correlations could arise if there is an adverse productivity shock, perhaps due to more stringent environmental regulation of one of the local employers. This would tend to reduce output, employment, and wages, and raise observed employment concentration. Again, the relationship is not causal but correlational with the unmeasured productivity shock.

Even where there are too few employers bidding for a set of potential workers to ensure competitive wage-setting—classical monopsony—there may be little that competition policy can do. A coal company may have labor market power because it is a dominant employer in a rural county, but if that position is not due to acquisition of rival employers or exclusionary behavior, it is unlikely to violate antitrust law. This is the labor market analog of the Supreme Court’s opinion in *U.S. v. U.S. Steel Corp.* (1920) that “the law does not make mere size an offense.”

Moreover, the term “monopsony,” as it is generally used among labor economists, is not reserved for situations with too few employers to be competitive. Rather, the monopsony label often is applied to many deviations from a perfectly competitive outcome that are unrelated to the number or concentration of employers competing to hire from a pool of workers. These are associated with a wide range of frictions in labor markets, such as information failures, transactions and search costs, idiosyncratic match quality, unwillingness of workers to relocate, occupational licensing, and more (Council of Economic Advisors, 2016b). These frictions typically do not arise from a reduction in competition among firms, either through merger or coordinated conduct, although the frictions may lead to a reduction in competition among employers. Nor is it likely that many of these are created by coordinated conduct by firms to limit competition or by unilateral conduct to exclude or disadvantage rival employers. With some exceptions, antitrust enforcement generally is not an effective or appropriate tool to address problems such as these (Naidu & Posner, 2018; Rogers, 2018; Rose, 2019). But there may be other policies that, by addressing the underlying friction, could improve both the operation of labor markets and outcomes for workers.

These critiques in no way imply the absence of competition problems in labor markets. There surely are monopsonistic markets in which employers restrict hiring to keep wages low, and further consolidation in those markets will likely worsen the problem. For instance, Prager and Schmitt (2019) provide evidence of this in their study of hospital mergers. They show that, consistent with monopsony power, mergers that

---

21 Robinson (1932, p. 215) coined the term monopsony for “an individual buyer which will correspond to the name monopolist for the individual seller.” In modern usage, monopsony is applied to markets with few buyers (employers).
substantially increased concentration in local hospital markets reduced wage growth by 1.0% to 1.5% per year for specialized health-care workers (pharmacists and nurses) and skilled non-health workers (e.g., hospital administration), while low-skill and unskilled workers appear unaffected. Other examples of anticompetitive practices in some current labor markets include wage-fixing and “no-poach” cases, in which employers agree not to recruit from or hire each other’s workers; the increase in noncompete clauses that restrict worker mobility, even for low-skill occupations (Starr, Prescott, & Bishara, 2019); and the explosion of occupational licensing laws that reduce both entry into occupations and mobility of workers in these occupations across markets (Kleiner, 2015; The White House, 2015; CEA, 2016b; Nunn, 2018). Union coverage has declined over the past several decades, and legal protections for workers, particularly for collective bargaining and class action litigation, have been eroded, tilting bargaining power toward employers (Council of Economic Advisers, 2016b).

Understanding the most significant causes of competitive problems in labor markets, as in product markets, is essential to identifying the most appropriate and effective policy interventions.

4. Concentration in the Digital Economy: How Should We Think About This Sector?

The apparent dominance of many of the large tech firms—and their prevalence in the social and political lives of so many—has generated levels of concern that seem to have crossed over to alarm. Calls to break up Google, Amazon, and Facebook, or to subject these and other companies to public utility style regulation, are ubiquitous (Yglesias, 2019). Antitrust investigations of some set of these firms have been announced by both federal enforcement agencies and a coalition of state attorneys general. Understanding the nexus of competition as it currently exists among firms is difficult. Even greater are the challenges of predicting the future contours of competition and credibly documenting that for a judge, which is required for competition policy enforcement in the United States, or designing a regulatory intervention to replace or restore competition.

---

22 One might ask why those mergers were allowed by the antitrust agencies. It appears that most of them were too small to be reported to federal authorities, and were not investigated by state antitrust enforcers. Wollman (2019) discusses rising “stealth consolidation,” acquisitions that are below the Hart Scott Rodino reporting thresholds, and hence are consummated without antitrust review.

This section describes some of the market nuances that are important for understanding concentration in the digital economy, the nature of competition, and what the implications are for policy.

Many firms that operate in the tech space appear to dominate their space in this ecosystem. Firms like Google, Facebook, Amazon, and Apple operate platform markets, in which the firm connects consumers with content providers, sellers, or advertisers who want to reach them. These markets tend to be characterized by strong network effects—many people want to be on the most popular platform, since that gives them the most others to interact with. In this case, the more popular the platform, the more new users it attracts. These can provide powerful incentives for firms to compete through some combination of better product offerings, user experiences, prices, and innovation, to attract customers to their platforms. The successful firms in most of these examples generally have done just that, delivering substantial value to consumers. Network effects are amplified when user-generated data improves the effectiveness of algorithms used to deliver value to both sides of the platform, permitting larger platforms to develop higher-value products and experiences, increasing users and user data in a positive feedback loop.

But the strength of these network effects can make these markets highly susceptible to “winner-take-all” or “winner-take-most” tipping toward the largest firm. This might entrench the large incumbent, making it difficult for entrants or other smaller competitors to gain users and build scale. Entrenched firms may see less need to provide consumers with innovative or high-value offerings. In these circumstances, competition for the market, rather than competition in the market, may be the primary constraint on incumbents. That is, a credible threat of entry and replacement by a new entrant may be the main limit on a dominant firm’s extraction of rents from consumers, unlike most conventional markets in which price and quality competition among existing firms generates value to consumers.

Second, it is important to recognize that “tech” is neither an industry nor a market. Business models of each of the large tech firms vary substantially, and even with the understanding that firms may share an emphasis on monetizing the value of consumer data, how firms do that and to what end may be quite different. Google has historically monetized the value of its search engine through sales of advertising

---

24 Katz and Sallet (2018) discuss the economics of platform (multi-sided) markets and propose the way courts should evaluate competition harms in such markets.

25 While we are accustomed to thinking about market power being exercised through higher prices, for many of these firms, users are enticed to a platform by “free” services. The platform profits by bundling these services with advertising and/or by collecting valuable data on users that is monetized. This is not a novel business model—radio stations, on-air television broadcasters, and many of the print media sources have for decades provided consumers with free or low-price access to content, paid for through advertising revenues. Payment cards (Visa, MasterCard, Amex) often provide transaction services to cardholders at a negative price (cardholder rewards such as points or cash back on purchases), paid for by higher merchant fees to process debit and credit card transactions.
delivered to the highest-value customers in response to their search terms. It dominates online search and search-based advertising, with global shares in the 70% to 80% range for desktop search and above 90% for mobile search. But is the market in which Google competes “all online search queries”? If the relevant buyers on the other side of the platform are advertisers, is the market search-based advertising, or online advertising, or all advertising? What is Google’s position in those larger markets?

It may be tempting to overstate the cleanness of market boundaries, as well as the protection offered by incumbency. For example, e-marketing firms have been reporting for several years that more consumers now start their product-based search queries from Amazon.com, rather than Google (Garcia, 2018). Data on consumer search and purchasing behavior on Amazon’s site is especially valuable in predicting what products consumers might buy, and how to increase purchase probabilities on Amazon.com. This may make Amazon a competitor not only to Google in search, but also to the third party sellers it matches to consumers in the Amazon Marketplace.

While the tech firms share some features—platforms that connect individuals and content providers who want to reach them online, generation of valuable data on the behavior of agents on both sides of the platform, business strategies that monetize those data—their individual business models and nature of consumer interactions vary widely. Innovation in this space has been an important driver of both consumer value and monetization of that value for the platform. Any policy intervention must navigate a complex set of sometimes conflicting objectives. For example, privacy protections may create a wedge between services consumers value and platform monetization of consumer data, or correct a failure that occurs when consumers do not fully understand what data firms are collecting and how they use it—or some of each. A number of recent reports to competition authorities and others provide thoughtful discussions of possible policy directions in this area.

5. Competition Policy: Has the Pendulum Swung Too Far to Under-Enforcement?

This brief argues caution in making sweeping inferences on the state of competition in the United States from highly aggregated statistics. But is there more we can glean from examining the state of competition-policy enforcement? U.S. competition


27 For example, competition authorities in the United Kingdom and EU have sponsored reports on the digital economy (Furman, 2019; Cremer et al., 2019); the U.K. Competition and Markets Authority commissioned an independent review of tech industry mergers (Argentesi et al., 2019); and the University of Chicago Stigler Center established a committee to report on market structure and antitrust for digital platforms (2019).
policy is a deterrence-based system. This recognizes the difficulty of detecting, investigating, and litigating all violations of competition policy, and instead seeks to deter companies from violating the antitrust laws by establishing clear policies and case law, and consequential penalties for firms that step over those lines. If enforcement becomes more lax, or penalties less certain or severe, deterrence is less effective and anticompetitive behavior may proliferate.

U.S. public antitrust enforcement operates in three broad areas. Section 1 of the Sherman Act prohibits “contracts, combinations, and conspiracies” in restraint of trade. This provides civil and criminal penalties for collusion among competitors (price-fixing, bid-rigging, market division, etc.) and restricts contracts found to be anticompetitive (e.g., prohibitions on intermediaries steering customers toward lower cost providers, or certain most favored nations clauses imposed by dominant firms). Section 2 of the Sherman Act restricts unilateral conduct by firms that monopolize or attempt to monopolize a market. Examples include the Microsoft antitrust case decided in 2001, and U.S. v. AT&T, which was settled in 1982 with the company’s breakup. Merger enforcement is governed by Sherman Act Section 1 and Clayton Act Section 7, with pre-notification of mergers above certain thresholds (roughly $90 million in 2019) required by the Hart Scott Rodino Act.

There are a number of reasons to believe that antitrust enforcement has become less vigorous over recent decades. Many observers suggest a decline in enforcement against anticompetitive conduct, pointing to examples like the dearth of Section 2 monopolization cases over the past 20 years, or the inability of the FTC to deter brand pharmaceutical firms from moving from one exclusionary tactic, like pay-for-delay of generic entry, to another, such as sham citizen petitions or denial of product to generic firms preparing an entry application to the Food and Drug Administration (FDA) (Feldman & Frondorf, 2016; Liu, 2017; Hemphill, 2006). Decisions by the enforcement agencies, particularly the Department of Justice (DOJ), undoubtedly play an important role in this outcome. But agency passivity is likely also due to much greater burdens of proof the courts have placed on plaintiffs across a wide range of anticompetitive behaviors. In some areas of antitrust enforcement, the courts now show great tolerance of behaviors that would have been considered per se illegal 50 years ago. Shapiro (2019, p. 80) terms this “the shrinking scope of the

---

28 The U.S. statutes also permit private antitrust enforcement, with treble damages if the plaintiff can prove anticompetitive harm. Private enforcement typically focuses on harm from collusion or exclusionary (monopolization) behavior, although a private plaintiff recently prevailed in divestiture it sought in a merger challenge, currently on appeal (Steves & Sons, Inc. v. Jeld-Wen, 2018). The courts have narrowed the scope for private antitrust enforcement over time, just as they have done with public enforcement.

29 The Antitrust Division issued guidelines for Section 2 enforcement during the waning days of the George W. Bush administration that were widely seen as affirming the Division’s abdication of enforcement against this conduct. See U.S. Department of Justice (2008). These were withdrawn as one of the first actions of Assistant Attorney General Christine Varney in 2009, but many observers note this was not followed by increased filing of Section 2 cases.
Sherman Act.” As a consequence, some problematic conduct has become almost unenforced against, if not unenforceable. Among these are predatory pricing and other predatory behavior; vertical restraints such as resale price maintenance and exclusive distribution contracts; and contracts that reference rivals such as most favored nations clauses (Baker, 2019; Shapiro, 2019). Exclusionary behaviors in most platform or two-sided markets may seem poised to join these categories in the wake of the Supreme Court’s decision in American Express (Ohio et al. v. American Express Co. et al., 2018).

There are also signs that merger enforcement has weakened. Some types of mergers have proven difficult for the agencies to prevail against in court. These include vertical mergers, where firms are related along a supply chain, such as the recent AT&T/Time Warner, Inc. merger, and potential competition mergers, where the parties are not significant active competitors with one another, as is common in much of the tech space. Mergers that fall below the Hart Scott Rodino notification thresholds, which have increased substantially over time, appear more likely to involve competitors and substantially less likely to attract enforcement attention (Wollman, 2019). Cunningham, Ederer, and Ma (2018) analyze mergers in the pharmaceutical space and find that firms are more likely to acquire and terminate competitive drug pipelines—what they call “killer acquisitions”—especially when they can do so below Hart Scott Rodino-reportable thresholds.

Second, the market structure threshold for challenges appears to have increased substantially over recent decades. The FTC periodically reports the fraction of merger investigations that resulted in an enforcement action (including litigated challenges, settlements, and abandonments). Kwoka (2017) analyzes FTC data for the 1996–2011 period, and reports enforcement rates binned by the number of “significant competitors” who would remain in a market were the merger allowed (roughly defined as the number of remaining firms with 10% or greater market share). The probability of a challenge if only one to four competitors would remain is above 50% over the entire period. In contrast, enforcement actions drop to zero for mergers with more than four competitors remaining by 2008–2011. And this is conditional on an investigation being opened, which is done only when staff have a reason to think the merger could be anticompetitive.

Third, this higher threshold for enforcement action is reflected in, and reinforced by, the evolution of the Horizontal Merger Guidelines over time. The HMGs offer guidance on the way the antitrust agencies approach merger investigations and challenge decisions. The first HMGs, issued by the DOJ in 1968, indicated the Division would challenge the acquisition of a 2% competitor by a 10% share incumbent in a highly

---

30 See the articles in the May 2018 Yale Law Journal “Collection: Unlocking Antitrust Enforcement” for the challenges and potential to bring cases in many of these areas under current case law.
concentrated market (CR4 above 75%), and its acquisition of a 4% share competitor in a moderately concentrated market. This reflected in part the prevailing hostility of courts to horizontal mergers, even in markets with relatively low concentration, and pushed against that hostility to loosen standards at least a bit. Thus, the infamous 1966 Von’s Grocery merger would not have triggered a challenge under the 1968 guidelines. The revision to the HMGs put in place by Attorney General Bill Baxter in 1982 ratcheted up the threshold for challenge and calibrated them to HHIs. The 1982 guidelines suggested the government was “likely to challenge” those mergers that increased HHI by more than 100 points *and* to a level above 1800 (on a 0–10,000 HHI scale); neither the acquisition of a 2% competitor nor one of a 4% competitor by a 10% share firm would trigger a challenge under the 1982 guidelines, regardless of other firms’ shares (Hovencamp & Shapiro, 2018). By 2010, the guidelines jointly issued by the DOJ and FTC had increased the threshold for highly concentrated markets from 1800 to 2500, raised the threshold for mergers that are “presumed to be likely to enhance market power” (and therefore likely to be challenged) to an increase of more than 200 points in a highly concentrated market, and stated that mergers leading to an increase in HHI of less than 100 “are unlikely to have adverse competitive effects and ordinarily require no further analysis” (U.S. Department of Justice, 2010, p. 19).

The evolution of the guidelines reflects a combination of changes adopted by the enforcement agencies, in part reflecting changed economic assessment of the likely costs and benefits of mergers, and in part a feedback loop between agency practice and court decisions that has ratcheted up the standards applied to merger challenges. The HMGs both inform courts and are informed by court decisions. There is growing concern that the structural presumption of harm for horizontal mergers has been excessively weakened over the past 40 years, both in terms of the level and changes of concentration at which it is applied and the deference given to it by the courts (Baker, 2019; Hovenkamp & Shapiro, 2018; Shapiro, 2019; Nocke & Whinston, 2019). This appears to reflect in part misplaced concern about the relative costs of overenforcement versus underenforcement, encouraged by an erroneous interpretation of the “Chicago School’s” theory that unfettered markets are competitive markets as an empirical fact. The outcome may have been encouraged by some hubris in the economics profession with regard to being able to measure with precision any potentially problematic effects arising from either mergers or anticompetitive conduct, encouraging courts to expect detailed quantitative evidence on competitive effects. This has led to greater roles for complex analyses

---

31 The Supreme Court, in *U.S. v. Von’s Grocery Co.* (1966), upheld the FTC’s challenge of a supermarket merger by the third and sixth largest firms in the Los Angeles market, that would have led to a combined market share of 7.5%.

by dueling economic experts and a movement away from structural presumptions. It is far from clear that this approach leads to better decision-making by lay judges untrained in economics and unaccustomed to antitrust cases on their dockets.

Fourth, strained agency resources likely contribute to underenforcement. Budgets are not keeping pace with challenges to competition. The budgets of the DOJ Antitrust Division and the FTC have increased only modestly in real terms over the past 20 years, while merger activity has skyrocketed, as shown in Figure 5. While the number of mergers in the U.S. economy has increased five- to seven-fold since 1985, the Antitrust Division budget has increased less than 60% over the same period and has declined in real dollars over the past decade. Budget-constrained staffing and the adverse effects of various federal hiring freezes limit the number of investigations that can be carried out simultaneously. Compensation for the professional staff—lawyers and Ph.D. economists—has been falling further behind private sector starting salaries, with likely consequences for both hiring and retention. Moreover, the rise of multi-billion-dollar megamergers, for which the cost of antitrust clearance is a small fraction of the total deal costs, creates significant asymmetries between the government’s available resources and the litigation teams the merging parties can and do assemble on the other side. In fiscal 2017, for example, 255 of the 1,992 Hart Scott Rodino merger notifications involved transactions in excess of $1 billion.

**Figure 4: DOJ Antitrust Division Budget Compared to Merger Activity, 1985-2017**

![Graph showing the comparison between DOJ Antitrust Division budget and merger activity from 1985 to 2017.](https://imaa-institute.org/mergers-and-acquisitions-statistics/)

*Source: Author’s calculations from Department of Justice (2018); All-Urban Consumer Price Index; IMAA Institute, https://imaa-institute.org/mergers-and-acquisitions-statistics/*
The Hart Scott Rodino Act puts merger review on a tight time clock; if the agencies don’t have the lawyers and economists to review an acquisition within those time limits, the parties are free to close on the transaction. One might expect the effects of budget constraints to show up as declining investigation and challenge probabilities during merger waves. These pressures also may impede conduct investigations, which often require considerable input of staff time to obtain and review documents and data, develop theories of harm, and assess the evidence. In a resource-constrained environment, there may be strong incentives to pull staff off a conduct investigation proceeding on an agency’s timetable to investigate a merger that will otherwise be consummated in 30 days.

5.1 Is U.S. Competition Policy Enforcement Lagging That in Other Developed Economies?

Resources aside, approaches to merger analysis, including consideration of theories of harm, the use of economic analysis, and application of a consumer welfare standard are broadly similar between the U.S. DOJ and FTC and their counterparts in the European Commission’s Directorate General for Competition (DG Comp) and the U.K. Competition and Markets Authority (CMA). It is uncommon for these authorities to reach substantially different outcomes in investigations of mergers that affect multiple jurisdictions, unless the competitive impact across those jurisdictions differs substantially due to different fact sets. There are some significant differences in process. For example, in the European Commission, merger enforcement is an administrative, not judicial process, so DG Comp is not required to convince a lay judge of the merits of its case in order to block a merger. 33 Non-litigation based processes like this could at the margin change the evidentiary standard, but this generally has not opened substantial gaps between U.S. and E.U. jurisdictions in merger enforcement. 34 It is noteworthy that in the tech space, for example, both the U.K.’s Office of Fair Trading (precursor to the CMA) and the FTC cleared Facebook’s acquisition of Instagram, now frequently cited as an acquisition by Facebook to co-opt a competitive threat (Baker, 2019); some decisions may look different with 20:20 hindsight. Given the different enforcement environments, it is interesting to speculate whether the observed convergence reflects a common emphasis on economic quantification, and its inherent limits in defining such a challenging counterfactual.

33 In principle, European Commission competition authority decisions can be appealed to the judiciary, but the delays this generates are generally seen as so costly that its merger decisions are rarely appealed. The FTC has a similar administrative process with internal Administrative Law Judges who could hear an FTC challenge, but the FTC increasingly enforces its merger actions through Preliminary Injunction hearings before a federal District Court judge.

34 Some argue that non-price harms, such as harms to innovation, are easier to act upon in European Commission merger investigations. This was argued in the Dow-DuPont merger, which the European Commission cleared with required divestiture of DuPont’s global research and development assets to preserve innovation competition that the United States did not insist upon. Officials from the DOJ and DG Comp disagreed with this characterization of the reason for that divergence (Guniganti, 2017).
Differences between the European Union and the United States are more significant in conduct enforcement—what the United States would term unilateral action or monopolization and the European Commission terms “antitrust” (Sokol, 2017). The European Commission operates with an “abuse of dominance” standard that is broader than the U.S. Section 2 monopolization standard, enabling the commission to enforce against behaviors that would not be a violation of U.S. law. Knowledgeable and reasonable voices disagree over whether some European Commission sanctions against U.S. tech companies like Amazon, Apple, and Google reflect more assertive antitrust enforcement or action against legitimate competitive conduct (Sokol, 2017; Shapiro, 2019). But it also may be easier for DG Comp to meet standards of proof under its standard than it would be for U.S. enforcers to invoke Section 2, and increasingly so given the U.S. Supreme Court’s higher evidentiary thresholds for Section 2 cases, even for exclusionary behavior that could be considered illegal under both regimes.

6. Restoring Competition Policy for a 21st Century Economy: What Are the Most Promising Directions?

There are myriad proposals for how to address concerns about increased market concentration or decreased competition in one or more sectors of the economy. These range from modest tweaks to the current system to dramatic overhauls that would change the objectives of antitrust as well as the processes. In the tech sector, proposals run the gamut from setting interoperability standards to requiring data exchange, imposing codes of conduct to limit exclusionary behavior, suing to unwind past mergers, breaking up large tech firms in the model of AT&T’s 1982 settlement, or creating public utility style regulation of platforms. Some of these offer the promise of more effective competition policy; others may reflect a naiveté about the constraints of antitrust enforcement or efficacy of regulation, or could do more harm than good. I offer below a number of promising directions for reform, distinguishing between what invigorated enforcement agencies could do and what is likely to require legislative intervention.

6.1 Increase Enforcement Agency Resources

Substantially increasing the Antitrust Division and FTC budgets is a straightforward and direct remedy to the stagnant resources enforcers have had to work with amid an increase in both the number and scale of merger activity. This can be made budget neutral by restructuring Hart Scott Rodino filing fees to move with the scale

35 Some of this work is by scholars with deep roots in legal antitrust scholarship or industrial organization (or both), often with enforcement experience at the DOJ or FTC (or both). See Baker, Sallet, and Scott Morton (2018) and the articles in “Collection: Unlocking Antitrust Enforcement”; Baker (2019); Sallet and Scott Morton (2018); Shapiro (2018, 2019).
of the proposed transaction, as proposed in recent legislation co-sponsored by Senators Amy Klobuchar (D-Minnesota) and Chuck Grassley (R-Iowa). Exit of career staff appears to be particularly high in the Antitrust Division under the present administration, and the DOJ has maintained a partial hiring freeze,\(^\text{36}\) which likely will make it necessary to invest significantly in rebuilding staff and capabilities. Restoring competitive pay scales, particularly for Ph.D. economists in the enforcement agencies, could help tremendously with that effort.\(^\text{37}\)

Of course, increased resources will lead to more enforcement activity only if the agency leadership is committed to vigorous enforcement. Increased agency budgets could usefully be accompanied by an earmark for review and assessment of past enforcement decisions. The FTC has a small research group, and FTC economists have had a merger retrospective research program for some time. DOJ could be encouraged or required to establish a similar program, and both agencies given authority to compel limited data production from parties to past investigations. These studies could provide information on whether anticipated outcomes were realized in markets, leading to improved agency decision-making, and possibly highlight when enforcement was too lax.

### 6.2 Empower Agencies to Pursue More Assertive Enforcement Profiles

While courts have increasingly narrowed the range of antitrust violations and increased the difficulty of winning cases for plaintiffs, there is both economic and legal support for more vigorous enforcement by the agencies. For example, the Baker et al. (2018) collection of articles in the *Yale Law Journal* highlights ways enforcers could bring and win more cases within the constraints of current case law, across a range of anticompetitive activity.

Innovation is not new in the agencies; staff adapt their understanding of the competitive dynamics of markets to new realities, revising their theories of harm and testing those against new evidence. Promoting a robust interaction between agency economists and academic researchers can be important in developing new theories and tools for enforcement and encouraging academic research to educate and validate these for enforcers and the courts. For example, bargaining leverage models were developed by the FTC to measure the anticompetitive effect of hospital mergers. This theory of harm has been adapted to a number of industry settings by both agencies and the Federal Communication Commission (FCC), which shares

---

\(^{36}\) The Division declines to make employment data available, so this is based on incomplete press reports of career staff exits.

\(^{37}\) Government salaries for the agencies’ Ph.D. economists lag shockingly behind academic and private sector salaries for new Ph.D.s. Requiring the Office of Personnel Management (OPM) to align their compensation with that for Ph.D. economists in financial agencies like the Treasury’s Office of Financial Research or the Federal Reserve Board would be a start. This could be done in conjunction with establishing a new economist employment category for federal hiring that would require a Ph.D., something OPM has vigorously resisted.
responsibility with the DOJ for many telecomm mergers, and an increasing academic literature illustrates its predictive power.

A number of avenues for invigorated enforcement are described below.

(i) More quickly embrace new economic models and new understandings of competitive dynamics.

In some cases where an agency’s decision not to challenge a merger has generated ex-post regret, it seems in part due to evolving understanding of the nature of competition in a particular market. More quickly recognizing and adapting to these new understandings could involve bringing cases that incur more litigation risk, but also with greater potential to extend the protection of competition policy. Early applications of new theories of harm may pose particular challenges for sorting out how to explain effectively the theory and evidence to a judge. That may be part of the reason the judge in the AT&T/Time-Warner merger litigation seems to have struggled to understand or accept the basics of the bargaining leverage framework (U.S. v. AT&T Inc., 2018). This is a price the agencies should be willing to pay for better grounded and more effective enforcement actions. Targets for this might include challenges to: vertical mergers (those between a firm and its supplier, or a firm and its distributor); mergers between competing employers that reduce competition for workers, or more generally buy-side mergers with the potential to harm upstream sellers (Hemphill and Rose, 2018); mergers that increase the probability of tacit collusion among firms (Baker et al., 2019); and mergers that harm innovation competition, particularly between firms without many current product overlaps but that are spurs to each others’ innovation activities. In some cases, this may move enforcers away from readily quantifiable harms, like increased price, to more qualitative harms, like diminished innovation competition. This would be a significant deviation from a 40-year trend toward increasing quantification of economic arguments in merger litigation, but it could be vital to agencies blocking important anticompetitive outcomes.

(ii) Adopt lower thresholds to determine merger challenges.

The HMGs say that agencies are likely to challenge further mergers in highly concentrated markets, but do not preclude a challenge of mergers below the 2500 HHI threshold for highly concentrated markets. Agencies could increase enforcement actions, starting with the moderately concentrated range, from roughly none today. This could be done without revision to the merger guidelines. But given the deference the courts give to the HMG, it may be better to issue a revision that acknowledges economic evidence that shows unilateral harms at levels below the current 2500 HHI cutoff and, in some markets, increased risk of tacit collusion.
(iii) Be less willing to settle problematic mergers.

The legal system prefers settlements to litigation. This is especially problematic in merger enforcement. If agencies identify a merger as anticompetitive, any negotiated settlement risks adverse effects from asymmetric information. That is, firms have much better information than does the DOJ or the FTC on what remedies will minimally constrain their ability to profit from the merger and will agree to remedies that tilt the outcome in their favor. Failure to recognize this is particularly dangerous in conduct remedy negotiations, in which firms agree to behavioral restrictions that are supposed to limit their ability to act on merger-created incentives to reduce competition. It is also problematic in so-called structural remedies, which involve divestitures of some assets. Even if the divested assets remain in business, many divestiture remedies, particularly partial or piecemeal ones, fail to restore fully the vigor of competition lost by the merger. And even where that is successful, if enforcers clear mergers with divestitures in selected markets that simply reduce any post-merger concentration level to 2499 or below in all markets affected by a merger, they may find over time that all markets converge to just below the threshold of high concentration.

If a merger is anticompetitive in more than a de minimis number of markets, agencies could sue to block the entire merger based on those affected markets, avoiding the potential for a failed divestiture and preserving competition. Antitrust is a deterrence system. If mergers that create anticompetitive harm are challenged rather than settled piecemeal, firms considering a merger that harms competition in some set of markets may be more reluctant to gamble on clearance or to face litigation.

The Antitrust Division recently took a significant step back from accepting behavioral remedies to vertical mergers. If this leads to challenges rather than settlements or clearances, it will be a welcome improvement: If a merger creates the incentive and ability to exclude rivals or raise their costs, enumerating in a consent decree a list of behaviors the firm agrees not to engage in postmerger is unlikely to eliminate the threat to competition. Antitrust agencies and courts are not regulators. If the firm has agreed not to take action A, which is profitable, it has every incentive postmerger to find action A’, which was not ruled out in the decree.

(v) Consider adoption of new frameworks for assessment of vertical mergers like AT&T/Time Warner, Inc. or CVS/Aetna.

The DOJ had not litigated a vertical merger challenge in 40 years when the Division sued to block AT&T/Time Warner. That case faced a number of challenges, some of which arose from disagreement over how to analyze efficiencies and exclusionary

38 The recent DOJ settlement proposed for the Sprint-T-Mobile merger is viewed by many observers as an exemplar of a remedy that in no way resolves the anticompetitive harm of the merger. The state attorneys general who are suing to block this merger in federal district court apparently agree.
incentives in a vertical combination. There is broad consensus that the DOJ’s Non-Horizontal Merger Guidelines, issued in 1984, are badly out of sync with economic understanding of vertical mergers and potential exclusionary behavior, and provide little helpful guidance to agency staff or the courts. While vertical guidance may be difficult to generalize, given the deference the courts have shown for the Horizontal Merger Guidelines, it may be worthwhile to enunciate a set of principles to guide challenges in vertical mergers. Baker et al. (2019) suggest principles that could ground such an effort.

(vi) Develop tough standards for efficiency defenses

The agencies consider in their assessment of mergers whether credible merger-specific efficiencies would sufficiently lower costs so as to offset any upward pricing pressure from a merger of competitors. Agency economists and financial analysts evaluate these claims through a skeptical, but sophisticated, analytic lens. Efficiency defenses are harder to adjudicate in court given the complex evaluations needed to assess most efficiency claims. The Supreme Court has yet to accept efficiencies as a defense against an anticompetitive merger, although this may be primarily due to how long it has been since a merger case reached the Supreme Court. Lower courts have been moving in that direction and there is reason to think the current Supreme Court may be sympathetic to that defense. There is substantial danger that court rulings sympathetic to firms’ claims of efficiencies could give companies a path to consummate almost any anticompetitive merger. Given how little economic evidence exists to support ex-post efficiency gains from most mergers, it would be appropriate for the agencies to clarify and toughen the standards for when, if ever, and which efficiencies could be appropriately weighed to defend an otherwise anticompetitive merger.

6.3 Consider Legislation to Re-Set Presumptions and Burdens of Proof.

Even if the DOJ and FTC adopt a more vigorous enforcement profile, the roadblocks created by case law over the past 40 years and an increasingly conservative judiciary that has been educated to accept the Chicago School’s skepticism of antitrust enforcement will be significant impediments to success. We may not have the luxury of 40 more years to attempt to gradually nudge the antitrust pendulum back. More timely progress likely requires legislation that re-establishes Congressional intent to enforce against a range of anticompetitive behaviors.

In mergers, this legislation may be most productively directed toward tightening the structural presumption, which benefits firms, enforcement agencies, courts, and the public, by making enforcement more transparent and redefining the expectations.

39 See Judge (now Justice) Kavanaugh’s D.C. Circuit Court of Appeals’ dissent in the DOJ’s suit to block the health insurer merger between Anthem and Cigna.
around certain burdens of proof. Progress on some of the thorniest antitrust enforcement challenges—potential competition and vertical mergers, predation, exclusionary conduct, and perhaps expectations for burdens in multi-sided markets—is likely to make little headway absent legislative intervention.

While it may be tempting to add additional objectives into the legislation (some have suggested a “public interest” standard for mergers, for example), experience with agencies or jurisdictions that have such expansive sets of objectives should give one considerable pause. The FCC has such a public interest standard; its merger investigations are sometimes characterized as holiday shopping expeditions for affected interests, as rent-seeking gives rise to “trades” of benefits in exchange for not opposing a merger. Similar concerns arise in proceedings before South Africa’s and China’s competition authorities. The current “consumer welfare” standard, properly understood to mean “trading partner welfare,” has been a serviceable standard, and is fully compatible with enforcement against non-price harms to customers, such as reduced quality, service, innovation, or other terms of trade, or upstream harms to seller, like reduced input prices due to a reduction of competition among buyers or employers.

Should policymakers want to break up tech firms, or unwind large numbers of past mergers, as Senator Elizabeth Warren (D-Massachusetts) has proposed, they are likely to find it difficult to do so through the courts. Those who would point to the 1982 AT&T disintegration as an example should recall that the DOJ began two monopolization investigations in the late 1960s—filing suits against IBM in 1969 and against AT&T in 1974—and reaching outcomes in each only in 1982, with abandonment of the IBM case after years of litigation and settlement of AT&T. And those took place in a legal environment that was much more conducive to conduct cases than is today’s environment. Even unwinding a completed merger may be extraordinarily difficult, as the government may have to prove not that the two parts of the firm would compete today, but likely that they were at least potential competitors or competitive threats at the time of the acquisition. There are many open questions about whether restructuring tech is wise or beneficial; there are even more about the efficacy of likely court challenges under current antitrust law.

6.4 Is There a Role for Regulation?

Some have advocated regulation of digital platform companies, rather than attempting to break them up. In this vein, antitrust and regulation could be seen as alternatives means to address a common problem—market power. Antitrust is

---

40 Warren’s platform proposes to “appoint enforcers” committed to unwinding a number of specific mergers in the tech space, including Facebook/Instagram and WhatsApp; Google/Doubleclick, Waze, and Nest; Amazon/WholeFoods and Zappos. It also proposes to require companies to divest digital platforms to a separately owned firm that does not participate on either side of the market.
the prophylactic, intended to prevent market power from arising, at least through anticompetitive means. Regulation is the curative, imposed on a firm or sector with market power to limit its exercise. If a large digital company has already acquired market power, is regulation the appropriate response?

Regulation is far from a cure-all, and there is considerable evidence from the history of economic regulation to suggest that in many cases, the remedy may be worse than the disease. A rich literature in regulatory economics warns of the costs of regulatory rent-seeking that raises entry barriers, facilitates legal collusion among firms to raise prices, and may impede innovation and dynamic efficiency (Joskow & Rose, 1989; Rose, 2014a, 2014b; and the references therein). Asymmetric information between regulators and firms creates opportunities ripe for rent-seeking behavior and regulatory distortions that may interfere with efficient operation (Laffont & Tirole, 1993). These failures appear particularly likely in highly dynamic sectors, in which innovation is important. Interest group capture tends to be most prominent among sector regulators; the breadth of antitrust agency authority has long been held to be one of the defenses against such capture of antitrust processes.

If regulation is desired as a policy response to unavoidable market power, the most promising direction is likely to be interventions such as the “light touch” approach outlined in the Furman Report (2019), perhaps focused on creating interoperability and data portability that facilitate entry and competition. Giving a regulator limited authority may help to reduce capture by the regulated firms. Replicating the academic expertise of leadership in a number of the U.K. authorities, rather than the political connections that are common in U.S. regulatory agencies, could be another way to reduce capture and improve decision-making, though this is not a model that has been used for many U.S. regulatory appointments in recent decades.

In some settings, it may be current regulation that impedes competition. Regulation that is motivated by rent-seeking or that is misdirected can lead to a variety of adverse market outcomes, as is likely for much of the growth in occupational licensing that impedes labor market entry. Mitigating these effects suggests reducing or eliminating licensing requirements where they serve little purpose in protecting consumers, such as licensing of florists, interior designers, or beekeepers. In cases where some oversight may be desired to protect consumers from their inability to discern the quality of providers—say in health-care provision or plumbing or electrical work—designing programs with minimally sufficient criteria to ensure appropriate training can achieve desired outcomes at lower costs. Another example of rent-seeking regulation are the state health-care laws passed to insulate hospitals from FTC merger review, allowing consolidation that reduces competition and raises prices. Antitrust agencies serve important competition-advocacy roles in settings such as these, but it requires politicians to put competition goals ahead of rent-seeking by important, and often well-funded, constituent interests.
7. Conclusion

In conclusion, as Baker (2019), Shapiro (2019), and many others have argued, government has likely retreated too far from the role it assumed almost 130 years ago with the passage of the Sherman Antitrust Act to ensure open, fair, and competitive markets. Rebalancing competition law to invigorate enforcement will require a combination of agency action and legislative intervention. But some competition problems may not be actionable through antitrust enforcement. In these cases, recognizing that both markets and regulation are imperfect is essential to determining whether intervention is likely to improve outcomes, and to designing effective policy in those cases.
Appendix A1: Industry Classification

For many classifications, the errors introduced by using three- or four-digit NAICS codes to define an industry can be substantial. Consider Food Manufacturing: In the multinational KLEMS ([K]apital-Labor-Energy-Materials-Services) databases, this would fall under KLEMS 10-12, which includes all manufacturing of “Food products, beverages and tobacco.” That level of aggregation combines a large number of industries that are neither rivals in consumer choices nor similar in production techniques or assets. The three-digit NAICS industry 311, “Food Manufacturing,” illustrates the problem.

KLEMS 10-12: Food products, beverages and tobacco

NAICS three-digit: 311: Food Manufacturing, which includes among others

- 311111 Dog and Cat Food Manufacturing
- 311230 Breakfast Cereal Manufacturing
- 31135 Chocolate and Confectionery Manufacturing
- 311511 Fluid Milk Manufacturing
- 31161 Animal Slaughtering and Processing
- 3117 Seafood Product Preparation and Packaging
- 311942 Spice and Extract Manufacturing

It is immediately apparent that this three-digit NAICS aggregation combines products and firms that are not in the same market: While one could debate whether Hershey’s cocoa powder competes with Teuscher truffles for consumer purchases, no one would likely suggest it competes with Purina Cat Chow, Tyson’s chicken carcasses, or Stonyfield organic yogurt. But the problem is far from eliminated by moving to the four-digit NAICS level. Consider NAICS code 3112, Grain and Oilseed Milling. It includes the following products:

- 3112 Grain and Oilseed Milling
- 31121 Flour Milling and Malt Manufacturing
- 311211 Flour Milling
- 311212 Rice Milling
- 311213 Malt Manufacturing
- 31122 Starch and Vegetable Fats and Oils Manufacturing
- 311221 Wet Corn Milling
- 311224 Soybean and Other Oilseed Processing
- 311225 Fats and Oils Refining and Blending
- 31123 Breakfast Cereal Manufacturing
Flour, soybean oil, and high-fructose corn syrup are likely inputs in the production of breakfast cereals, and none of these products would appear to compete with one another. Nor are six-digit industries “correct.” The market for corn syrup, a ubiquitous sweetener, likely requires not only NAICS 311221 but also cane and beet sugar manufacturing (combined in NAICS code 31131; but in different four-digit SIC codes) and perhaps for some uses, honey processing (311999) or artificial sweeteners (325199 and 325998, within the chemical manufacturing sector of the NAICS codes).
Appendix A2: Firm Revenue Assignment

Studies generally do this one of two ways, based either on establishment-level information from the United States Economic Census or private data sources such as the NETS database (United States) or Orbis, or firm-level data, such as Compustat. Studies using firm-level data typically assign the firms’ entire reported global revenue to their primary reported industry category and home country. For larger firms that produce multiple products or operate across multiple markets, this can be quite misleading.

Consider three examples of multiproduct firms with segment reporting:

<table>
<thead>
<tr>
<th>3M (2015 10K)</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL FOR FIRM: Primary NAICS Code 322220: Cutting and coating paper and paperboard</td>
<td>$31.8B</td>
</tr>
<tr>
<td>Industrial: Tapes, coated, nonwoven and bonded abrasives, adhesives, advanced ceramics, sealants, specialty materials, filtration products, closure systems for personal hygiene products, acoustic systems products, automotive components, abrasion-resistant films, structural adhesives and paint finishing and detailing products</td>
<td>11.0B</td>
</tr>
<tr>
<td>Safety and Graphics: Personal protection products, traffic safety and security products, commercial graphics systems, commercial cleaning and protection products, floor matting, and roofing granules for asphalt shingles</td>
<td>5.7B</td>
</tr>
<tr>
<td>Electronics and Energy: Optical films solutions for electronic displays, packaging and interconnection devices, insulating and splicing solutions for the electronics, telecommunications and electrical industries, touch screens and touch monitors, renewable energy component solutions, and infrastructure protection products</td>
<td>5.6B</td>
</tr>
<tr>
<td>Health Care: Medical and surgical supplies, skin health and infection prevention products, drug delivery systems, dental and orthodontic products, health information systems and food safety products</td>
<td>5.6B</td>
</tr>
<tr>
<td>Consumer: Sponges, scouring pads, high-performance cloths, consumer and office tapes, repositionable notes, indexing systems, construction and home improvement products, home care products, protective material products, and consumer and office tapes and adhesives</td>
<td>4.5B</td>
</tr>
</tbody>
</table>

---

41 Some studies adjust global revenues for imports and exports to yield U.S. revenues; occasionally this is done at the firm level to yield U.S. sales, more often at the industry level to adjust concentration for import competition. In almost all studies that do this, the adjustments are based on aggregate import/export statistics by industry.
### Archer Daniels Midland (ADM; 2017 10-K)

<table>
<thead>
<tr>
<th>Category</th>
<th>Net Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TOTAL FOR FIRM:</strong> Primary NAICS code 31122: Starch and Vegetable Fats and Oil Processing</td>
<td>$62.3B</td>
</tr>
<tr>
<td>Agriculture Services:  Grain storage, transportation networks, food and feed ingredients, structured trade finance, flour milling</td>
<td>27.9B</td>
</tr>
<tr>
<td>Corn Processing:  Corn wet milling and dry milling, ethanol production, bioproducts, feed additives</td>
<td>9.5B</td>
</tr>
<tr>
<td>Oilseeds Processing: soy, canola, sunflower, etc. processing for food, feed, energy, and industrial products</td>
<td>22.2B</td>
</tr>
<tr>
<td>Wild Flavors and Specialty Ingredients:  mfg, sales, distn of natural flavor ingredients, flavor systems, natural colors, proteins, etc.</td>
<td>2.5B</td>
</tr>
<tr>
<td>Other:  primarily financial, futures and insurance</td>
<td>0.4B</td>
</tr>
</tbody>
</table>

### DuPont (2015 10-K)

<table>
<thead>
<tr>
<th>Category</th>
<th>Net Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TOTAL FOR FIRM:</strong> Primary NAICS code 3252: Resin, Synthetic Rubber, and Artificial Synthetic Fibers and Filaments Manufacturing</td>
<td>$35B</td>
</tr>
<tr>
<td>Agriculture:  seeds, crop protection chemicals (54% outside US)</td>
<td>11.3B</td>
</tr>
<tr>
<td>Performance Chemicals:  Titanium tech, chemicals &amp; fluoroproducts, Teflon, commodities</td>
<td>6.5B</td>
</tr>
<tr>
<td>Performance Materials:  Polymers, resins, elastomers (70% sales outside US)</td>
<td>6.2B</td>
</tr>
<tr>
<td>Safety &amp; Protection:  Personal and environment protection, incl. Kevlar, Nomex, Tyvek; homeland security consulting; solutions for construction, transportation, communication, etc.</td>
<td>3.9B</td>
</tr>
<tr>
<td>Nutrition &amp; Health:  specialty food ingredients, food nutrition, health and safety</td>
<td>3.5B</td>
</tr>
<tr>
<td>Electronics &amp; Communications:  for photovoltaics (PV), consumer electronics, displays and advanced printing</td>
<td>2.4B</td>
</tr>
<tr>
<td>Industrial Biosciences:  biobased products for animal nutrition, detergents, food manufacturing, ethanol production and industrial applications</td>
<td>1.3B</td>
</tr>
</tbody>
</table>
A decision to assign firm-level revenues to the reported primary NAICS/SIC code would substantially misrepresent each of these firm's activities, both in the primary market and in all others in which it operates. Even with the much-abbreviated descriptions of segments included in the tables above, it is clear that most reported segments are an agglomeration of many different products, markets, and sectors that are not competitive with each other from the standpoint of customers, and may not be close substitutes in terms of production facilities and technologies. Finally, the mix of U.S. and global sales reported for some of the segments highlights the danger in assuming all or most revenue is U.S. revenue.42

42 While some studies attempt to adjust for imports and exports using U.S. aggregate import/export shares by industry, there is no reason to think applying these high-level aggregates to firm-level data will produce accurate adjustments.
References


Concerns About Concentration


Robinson, J. (1932). The Economics of Imperfect Competition. Springer.


PART II

RISING FEDERAL DEBT AND SLOWING ECONOMIC PRODUCTIVITY

Fiscal Policy With High Debt and Low Interest Rates
William Gale

Can Innovation Policy Restore Inclusive Prosperity in America?
John Van Reenen
Fiscal Policy With High Debt and Low Interest Rates

AUTHOR
William Gale

ABSTRACT
Policymakers in the United States face a combination of high and rising federal debt and low current and projected interest rates on that debt. Rising future debt will reduce growth and impede efforts to enact new policy initiatives. Low interest rates reduce, but do not eliminate, these concerns. The federal fiscal outlook is unsustainable even with projected interest rates that remain below the growth rate for the next 30 years. Short-term policy responses should focus on investments that are preferably tax-financed rather than debt-financed. Most importantly, policymakers should enact a debt reduction plan that is gradually implemented over the medium- and long-term. This would avoid reducing aggregate demand significantly in the short-term and, if done well, could actually stimulate current consumption and production. It would stimulate growth in the long-term, provide fiscal insurance against higher interest rates or other adverse outcomes, give businesses and individuals clarity about future policy and time to adjust, and provide policymakers with assurance that they could consider new initiatives within a framework of sustainable fiscal policy.

* Brookings Institution and Tax Policy Center. This chapter was prepared for the Aspen Economic Strategy Group and is an adaptation and extension of Fiscal Therapy: Curing America’s Debt Addiction and Investing in the Future (Oxford 2019). I thank Aaron Krupkin and Victoria Johnson for outstanding assistance and Alan Auerbach, Doug Elmendorf, Amy Ganz, Marc Goldwein, Melissa Kearney, Maya MacGuineas, Emily Merola, Brian Riedl, and David Wessel for helpful comments.
1. Introduction

U.S. policymakers face a combination of high and rising federal debt and low current and projected interest rates on that debt. Figure 1 shows that the debt-to-GDP ratio is at its highest level in U.S. history, except for a few years around World War II, and that government net interest payments (the product of debt and the average interest rate) are currently at their average historical level as a share of the economy.

**Figure 1: Debt and Net Interest as a Share of GDP, 1940-2019**

Rising future debt will slowly but surely make it harder to grow our economy, boost our living standards, respond to wars or recessions, address social needs, and maintain our role as a global leader.

Lower interest rates reduce these concerns, holding other factors constant (such as economic growth rates). At the very least, low interest rates undermine claims that current debt levels will cause a financial crisis. More generally, low rates reduce the fiscal cost of debt accumulation. To the extent that low interest rates indicate a reduced marginal private return to capital, the opportunity cost of government borrowing falls, making it more attractive to pursue new government programs, particularly investments.

However, low interest rates are not a “get out of jail free” card. Although interest rates are low, seemingly every other major aspect of the fiscal situation is problematic. The
full-employment deficit is already high and is expected to remain at elevated levels in
the absence of policy changes; in the past, it spiked only on a temporary basis. Short-
term deficits will rise further if policymakers extend temporary tax and spending
provisions, as they have done repeatedly in the past. Over the longer term, even if
interest rates stay below the growth rate, interest payments will rise steadily to over
6% of the economy—as large as Social Security outlays—under standard assumptions.
And even if interest rates stay constant, interest payments will rise because the debt
is rising. Likewise, Social Security and health care outlays will continue to rise because
of the aging of the population. These three program areas—interest payments, Social
Security, and health care—will account for more than 100% of all federal spending
growth as a share of GDP. In contrast, federal investments in infrastructure, research
and development, and human capital are slated to decline. Meanwhile, several major
federal trust funds—including Social Security and Medicare—are slated to exhaust their
balances within the next 15 years; the budget is largely on autopilot, with mandatory
programs, which are not annually appropriated, accounting for an increasing share of
federal outlays over time; and the political system seems broken, with political leaders
unable to muster the cooperation and trust—or even the interest—that bipartisan fiscal
agreements typically require. But the longer we wait to make policy changes, the
larger and more abrupt those changes will need to be, unless interest rates stay at or
close to their current levels for the next 30 years.

Because of these considerations, low interest rates do not necessarily eliminate the
unsustainability of the long-term fiscal position of the U.S. government. This point is
explicitly recognized by those economists who argue most strongly for the salience
of low interest rates for policy choices (Blanchard, 2019a, 2019b; Elmendorf, 2019;
Elmendorf & Sheiner, 2017; Furman & Summers, 2019). Krugman (2019) and Furman
and Summers (2019) argue against expanding even the short-term deficit, except for
financing investments or fighting recessions.

Under what I view as standard assumptions, where future interest rates rise but
remain below the growth rate for the whole projection period, the debt-to-GDP ratio
is projected to rise more or less continually from its current level of 78% to 169% by
2049. To limit the debt-to-GDP ratio to 100% by 2049 would require permanent tax
increases or spending cuts starting in 2021 equal to 2.3% of GDP. If policy adjustments
are delayed to 2029, the required annual change would equal 3.2% of GDP. Even if
interest rates remain constant over the next 30 years, the debt would rise to 134% of

---

1 CBO (2019b, 2019c) projects exhaustion over the next 15 years for the trust funds for Social Security,
Disability Insurance, Medicare Part A, Highways, and Pension benefits and for the Guaranty Corporation Multi-
Employer Fund.

2 Other exacerbating factors include: The Fed’s efforts to unwind its portfolio of Treasury debt; rising debt
issued by the States; the low U.S. saving rate compared to many other countries; declines in foreigners’
williness to hold federal debt; efforts by other countries and leading corporations to develop new payment
systems that could threaten the dollar’s role as the world’s reserve currency. None of these factors is decisive,
in and of itself, but none of them helps the fiscal situation, either.
GDP and the required permanent policy adjustment to limit the debt-to-GDP ratio to 100% would be about 1.3% of GDP.

Finally, while lower interest rates improve the federal government’s overall fiscal stance—because it is a net borrower—they come with two additional caveats. First, we can certainly borrow more and consume more with low interest rates and not hurt future generations (who can in turn borrow more from later generations), but the optimality of this pattern falls apart if interest rates subsequently rise and we are left with higher interest rates on higher levels of debt. Second, low interest rates raise the present value of future spending obligations, like those for Social Security and Medicare. In the past, policymakers have chosen to pre-fund a certain share of these obligations. With lower interest rates, any level of pre-funding will be more difficult to achieve; (i.e., it will require higher taxes or lower spending than with higher interest rates). Policymakers will have to choose between imposing higher burdens to reach a given level of pre-funding or pre-funding these programs to a lesser extent than in the past.3

How should policymakers respond? They should not try to reduce the short-term deficit. That is not the problem; the long-term projection is. In addition, cutting current deficits would likely reduce aggregate demand, a change that monetary policy may be hard-pressed to offset, given low interest rates.

Policymakers should also enact new investment programs. We need more infrastructure, research and development, and human capital, even apart from the fiscal stance. I conclude that it would be preferable—based on fiscal and economic growth considerations—to fund these projects with taxes rather than deficits. But, if given the choice between deficit-financed investments and no investments, policymakers should choose the former.

In contrast, except for antirecession purposes, policymakers should not enact deficit-financed spending for non-investment programs, though they should embrace a broad definition of what constitutes an investment, to include programs that make people more productive by providing childcare, job training, and related items.

Finally, policymakers should enact in the near term a plan that is implemented on a gradual, phased-in basis and substantially reduces long-term deficits and debt from future projected values. This approach would avoid reducing current aggregate demand significantly. Indeed, if done well, it could boost current spending and production. It would help the economy in the long-term and thereby reduce burdens on members of future generations, many of whom will not be better off than their

---

3 When the Social Security or Medicare trust fund runs an annual surplus, the excess funds are invested in bonds at the Treasury. The interest rate that the Treasury Department pays to these programs depends on recent average yields on federal debt. As a result, lower interest rates reduce the returns that the trust funds receive and thus make it more costly to achieve a given level of pre-funding. (In a similar fashion, low rates of return make it more difficult for pension funds to finance future obligations.)
parents. It would provide some fiscal insurance against interest rate jumps or other adverse fiscal outcomes; with debt already at high levels relative to GDP and projected to rise, the budget is more sensitive to interest rate fluctuations now than it has been in the past.

A gradual phase-in would provide time for businesses, investors, and citizens to adjust their plans and would reduce political backlash. Finally, a debt reduction plan would give economic agents more certainty about future policy and offer policymakers assurance that they could undertake new initiatives within a framework of sustainable fiscal policy.

To move into specific examples, policymakers should consider three sets of policies.

- Some gradual, debt-reduction policies could stimulate consumption and production now. Enacting a consumption tax (value-added tax) whose rates rose gradually over time would stimulate current consumption as customers spent more today to avoid higher future prices. Likewise, introducing a carbon tax with rates that rise over time could stimulate current production, as producers choose to use more fossil fuels now while they are still relatively inexpensive.4 Both policies could generate significant long-term revenues.

- Some gradual, debt-reduction policies are needed in their own right. Making Social Security sustainable is one example. Another is boosting health-care coverage and reducing costs (by creating a public option on the exchanges, converting Medicare to a premium support plan, allowing Medicare to negotiate drug prices and formulary, and limiting the tax subsidy for health insurance for families with above median-cost plans).

- In addition to debt reduction, policymakers should initiate substantial new investment programs in infrastructure and research and development (1% of GDP) and in children, families, and human capital (another 1% of GDP). These changes can be financed by closing income tax loopholes and converting the corporate tax to a 25% cash flow levy.

All these policies could be phased in gradually, and together they would be sufficient to stabilize debt below 100% of GDP over the long-term. (They would reduce 2049 debt to 60% of GDP if initiated in 2021.)5 Even if policymakers adopt a different way to reduce long-term debt and make new investments in America, enacting a reasonable, gradual debt reduction plan would be a major improvement over the current situation.

---

4 Both the consumption tax and carbon tax could be accompanied by other policy changes that would offset the regressivity of the taxes. Phasing in a carbon taxes need not create a “Green Paradox” situation where imposing future carbon taxes raises overall greenhouse gas emissions. See Sinn (2012) and Williams (2016).

5 For more details on the effects of these policies, see Gale (2019).
These conclusions are guided by several overarching themes. First, the nation faces two intertwined problems: The rising, long-term debt profile and the way we tax and spend. Government spending is too oriented toward consumption relative to investment, the latter broadly defined to include human capital. Likewise, our tax system could be fairer and more efficient and could produce more revenue. Offering a debt reduction plan provides an opportunity to address simultaneously the debt problem and the structure and composition of taxes and spending.

Second, the economy is more important than the budget. Saving the budget but hurting the economy would be a Pyrrhic victory. This seems to be a particularly salient issue currently. Normally, in an economy with unemployment at a 60-year low and the full-employment deficit for several years projected to be at least 5% of GDP, the obvious prescription would be significant, fairly rapid fiscal consolidation—tax increases, spending cuts. With current low interest rates, low inflation, and concerns about weak growth even amidst remarkably accommodative monetary and fiscal policy, however, it would be prudent to make any fiscal adjustments gradually.

Third, historical patterns can inform the current fiscal situation. In many ways, we are in uncharted territory. We have never had to address the projected permanent imbalances between spending and taxes that we face now. In some ways, though, the closest historical antecedent to our current status occurred after World War II, when the United States faced even higher debt as a share of the economy than today, and even lower interest rates. But the cause of the debt (wartime spending) and the composition of government spending (low entitlement outlays) were very different then. In any case, the debt-to-GDP ratio fell rapidly as military spending fell and the nation essentially ran balanced primary budgets from 1947-1980. In contrast, we are currently projected to run permanent, substantial primary deficits.

Finally, apart from all the specific arguments, a broader view might be constructive. Although the interest rate on government debt has been less than the economic growth rate more often than not historically in the United States (Blanchard 2019a, 2019b) and in other countries (Mauro, Romeu, Binder, & Zaman, 2015), it appears to be a long-standing convention that governments do voluntarily run up their debt. In 2007, for example, before the financial crisis raised debt levels everywhere, only two OECD member countries (Greece and Italy) had general government net financial liabilities, relative to GDP, in excess of the current value for the United States (OECD, 2019). Keeping a lid on debt may be simply an outdated, prudish norm that does not apply to the economic situation facing the United States today. Alternatively, there might be very good reasons for this behavior—the desire to maintain “fiscal space,” a concern that high debt reduces growth and imposes burdens on future generations, etc.—and therefore some wisdom embedded in those established government practices.
The rest of the paper develops the points above in more detail. Section 2 reviews the fiscal outlook under varying assumptions about interest rates, to provide context and outline where we are headed. Section 3 explains why the projections are worrisome and discusses the economic effects of rising debt and deficits. Section 4 presents estimates of the fiscal gap, the size of policy changes needed to reach particular fiscal targets. Section 5 discusses in more detail how policymakers should respond. Section 6 concludes. (Appendix A provides details of the budget projections. Appendix B explains the flaws in three separate claims that debt does not matter.)

2. Where Are We Headed?

I develop 30-year budget projections under what I view as a continuation of “current policies.” (See Appendix A for details). The key, long-term assumptions relate to growth rates and interest rates. I follow the Congressional Budget Office (CBO) (2019c) in setting the nominal annual growth rate to average 4% after 2029 and the average nominal interest rate on government debt to rise gradually from 2.4% in 2019 to 3.9% in 2049. This 150-basis-point increase—which I call the “standard” scenario—keeps the interest rate below the growth rate throughout the entire period and is consistent with the effects of rising government debt (of the magnitude shown in the projections below that use this interest rate path) found in several studies.6

Because the fiscal outlook depends sensitively on interest rates and because financial market indicators currently imply lower future interest rates than the CBO does, I also consider a “flat” scenario, where the average interest rate on government debt is constant through 2049 at its 2019 value of 2.4%.7 To be clear, I regard this as an optimistic scenario, and I include it to highlight the effects of low interest rate projections on the fiscal outlook.

The budget outlook can be described in a series of graphs. Figure 2 shows that primary deficits (which exclude interest payments) will rise from 2.4% of GDP currently to 3.8% of GDP by 2029, then remain relatively constant through 2039, after which they fall to 2.7% of GDP by 2049. The main point of Figure 2 is that the federal budget is out of balance on a long-term basis, even ignoring interest rates and interest payments. In contrast, in the generation following World War II, the government ran primary surpluses more often than not and the primary budget averaged a deficit of just 0.1% of GDP from 1947-1980.

6 See Engen and Hubbard (2005); Gale and Orszag (2004); Gamber and Seliski (2019); Krishnamurthy and Jorgenson (2012); Laubach (2009); and Tedeschi (2019). These studies show that a 1 percentage point increase in federal debt as a share of GDP raises interest rates by 2 to 3 basis points.

7 Using interest rate forecasts based on current financial market data may not be appropriate in a budget projection because those forecasts presumably include a positive probability that some sort of budget deal is reached before the budget period ends, which is inconsistent with the assumptions in the budget projection.
The persistent primary deficits are best interpreted as the result of a long-term, sizable mismatch between what Americans want from their government and what they are willing to contribute, rather than as a “spending” problem or a “tax” problem. Figure 3 shows that non-interest spending is projected to be substantially higher throughout the next 30 years (20.8% of GDP) than its average since 1965 (18.2% of GDP). Notably, there are no new spending initiatives built into the projections, which simply show the playing out of commitments that political leaders made in the past. Given the aging of the population, it is virtually inevitable that government spending will rise. Projected revenues average 17.4% of GDP over the next 30 years, equal to their post-1965 average.

Figure 4 provides more details on the changing level and composition of non-interest spending, assuming there are no new spending programs for the next 30 years. Spending on Social Security and health care will rise, accounting for more than 100% of the increase in non-interest spending as a share of GDP. Outlays on all other non-interest categories will fall. Non-defense, discretionary spending includes most of the federal government’s investment projects in infrastructure, research, and education. Other mandatory spending contains most of the government’s safety net initiatives. These two categories and defense spending are all slated to fall by between 19% and 31% relative to GDP over the next 30 years.
Part II: Rising Federal Debt and Slowing Economic Productivity

Figure 3: Primary Deficit as a Share of GDP, 2019-2049

Source: Author’s calculation; Board of Trustees (2019); Boards of Trustees (2019); Congressional Budget Office (2019b, 2019c)

Figure 4: Non-Interest Spending Projections, 2019-2049

Source: Author’s calculation; Board of Trustees (2019); Boards of Trustees (2019); Congressional Budget Office (2019b, 2019c)
Figure 5 shows that the interest rate scenario makes an enormous difference to budget projections. With “standard” rates, interest payments rise from 1.8% of GDP in 2019 to 3.4% of GDP in 2029, and to 6.2% of GDP in 2049. If interest rates remain at today’s low levels, interest payments rise at a much slower pace, to 2.3% of GDP in 2029 and 3.1% of GDP in 2049. By comparison, interest payments averaged 1.8% of GDP from 1947 to 2019 and equaled 1.6% of GDP in 2018. In 1991, interest payments reached their historical peak of 3.2% of GDP. Thus, under the optimistic scenario (“flat” rates), interest payments are projected to rise almost to their historical maximum; under “standard” rates, interest payments will skyrocket to almost double their previous peak.

Figure 6 combines the primary deficit and interest payments to show projections for the federal deficit. As before, interest rate assumptions matter significantly. With standard interest rates, the deficit rises from 4.2% of GDP currently to 7.2% in 2029 and to 9% of GDP by 2049. With flat interest rates, the deficit still rises to about 6.1% of GDP by 2029 and 6.6% by 2040, after which it drops to 5.8% of GDP by 2049. The deficit falls somewhat in the out-years because of the very strong and restrictive assumptions about discretionary spending and other mandatory outlays described above.

Figure 7 shows historical and projected figures for debt as a share of GDP. With standard interest rates, debt rises from 78% of GDP in 2019 to 106% of GDP in 2029 and 169% of GDP by 2049. With flat rates, debt rises much more slowly, but it still rises inexorably and to all-time high levels. The debt rises from 78% of GDP in 2019 to 98% of GDP in 2029 and to 134% of GDP in 2049.

Several aspects of the debt projection are salient. First, the projected fiscal shortfall differs from those in the past in important ways. From the nation’s founding until about 1980, debt as a share of the economy rose only when we were at war or in recession, and it only rose temporarily. After the war or recession ended, debt fell rapidly. Starting in 1981, President Ronald Reagan’s tax cuts and defense-spending increases raised debt during peacetime prosperity. A series of largely bipartisan tax increases and budget deals from 1992 to 1997 helped turn persistent deficits into surpluses by the end of the century. Since 2000, tax cuts and spending increases under Presidents George W. Bush, Barack Obama, and Donald Trump, along with the Great Recession, greatly boosted current and projected levels of future debt.

The current economic and budget projections are different from those in the past. Relative to pre-1980 debt, current projected debt-to-GDP ratios are higher, and the trend is permanent. There is no war or recession that will end and let the budget adjust. Relative to the early 1980s, we now face a much higher initial debt level and the headwinds generated by demographics. In 1981, debt was only one-third as large as it is today relative to GDP, and the economy benefitted from the steady influx
Figure 5: Net Interest as a Share of GDP, 2019-2049

Source: Author’s calculation; Board of Trustees (2019); Boards of Trustees (2019); Congressional Budget Office (2019b, 2019c)

Figure 6: Deficit as a Share of GDP, 2019-2049

Source: Author’s calculation; Board of Trustees (2019); Boards of Trustees (2019); Congressional Budget Office (2019b, 2019c)
of baby boomers and women into the labor market. Now, boomers are retiring en masse and women’s labor force participation has plateaued.

The experience after World War II is particularly informative. After the war, the debt-to-GDP ratio fell more or less continually, from 106% of GDP in 1944 to about 26% of GDP in 1980. Part of the reason was low interest rates and strong growth, but another factor was that primary deficits were quite small (averaging just 0.1% of GDP from 1947 to 1980). Currently, however, as discussed above, primary deficits are projected to average 3.4% of GDP over the next 30 years, even under optimistic assumptions (Figure 2). While we don’t need to cut the debt to 1980 levels, we do need to stabilize it at a reasonable amount. That will require significant reductions in the primary deficit.

Second, long-term projections are sometimes dismissed on the grounds that they are subject to substantial uncertainty (Krugman, 2012). But it is not necessary to focus on the long-term to see the fiscal imbalance. Figure 8 shows that the current full-employment deficit is already high at almost 5% and will remain high—and actually rise—over the next decade, assuming that policymakers extend temporary tax and spending provisions, as they have done in the past (see Appendix A). Historically, full-employment deficits have been much lower (averaging 2.8% of GDP since 1965) and only spiked temporarily.

Another short-term indicator of concern is that, according to the International Monetary Fund (IMF), the United States is one of only a handful of advanced countries projected to experience a rise in their debt-to-GDP ratio over the next 5 years and indeed will experience the largest increase among those countries (International Monetary Fund, 2019).\(^8\)

Third, it is worth emphasizing that the projections above are based on relatively optimistic economic and policy assumptions. The economy grows steadily; interest rates stay below the economic growth rate; there are no unusual or deep recessions; and climate change does not impose any extra burden on the economy. Defense spending grows only with inflation. There are no new wars. There are no new, major spending initiatives, and domestic spending other than for Social Security and health care falls significantly relative to the size of the economy. One way to show how optimistic the assumptions are is to note that, under plausible alternative assumptions, the CBO (2019c) estimates the debt-to-GDP ratio would be 219%, compared to 169% in the estimates here.

---

8 Short-term projections are based on officially announced budgets, adjusted for differences between the national authorities and the IMF staff regarding macroeconomic assumptions. The medium-term fiscal projections incorporate policy measures that are judged by the IMF staff as likely to be implemented.
Figure 7: Debt-to-GDP, 1790-2049

Source: Author’s calculation; Board of Trustees (2019); Boards of Trustees (2019); Congressional Budget Office (2019b, 2019c)

Figure 8: Full-Employment Deficit as a Share of GDP, 1965-2029

Source: Author’s calculation; Board of Trustees (2019); Boards of Trustees (2019); Congressional Budget Office (2019b, 2019c)
3. Effects of Rising Debt and Deficits

If left unaddressed, rising deficits and debt will cause significant, long-term economic problems, curtailing growth and limiting the rise of living standards for our children and grandchildren. They will also hamper the government’s ability to address other issues and will reduce America's global standing. Despite public controversy about fiscal policy, there is a well-established consensus—even with current projections keeping the government interest rate \(r\) below the economic growth rate \(g\) and even among those who advocate not addressing the long-term fiscal situation now—that following our current fiscal path would do long-term damage to the economy and is unsustainable.9

3.1 Debt and the Economy

Not all debt is bad. As Alexander Hamilton explained in the 1790s, debt helps the government establish credit and trade with other nations. It gives investors a safe and liquid asset, provided the government stays solvent.10 It helps nations finance their responses to emergencies, such as recessions or wars, and it helps finance investments in people or projects that will raise future living standards. And, of course, deficits can provide a boost when the economy falls into recession.

In short, the effects of fiscal policies on the economy depend not only on the timing and size of the deficits but also on the specific policies that generated those deficits. The concern about the projected long-term debt build-up, then, is not just about the debt, per se; it is also about the way we are taxing and spending. The path we are on produces a rising debt-to-GDP ratio that essentially is financing increased transfer payments to the elderly (Figure 4, Figure 7). Meanwhile, public investment in infrastructure, scientific research, and human capital are projected to decline as a share of GDP, as are safety net expenditures that can help low-income families lead more productive lives.

Following this path will reduce future national income. If the government borrows to provide a tax cut or spending benefit, government saving falls by the full amount of the borrowing, while recipients save some (but typically not all) of the tax cut or the

---

9 See Blanchard (2019a, 2019b); Furman and Summers (2019); Elmendorf (2019); Elmendorf and Sheiner (2017). All conventional economic models suggest that high and rising debt-to-GDP ratios will hamper long-term growth. A 2013 survey of leading academic economists of varying political affiliations asked for reactions to this statement: “Sustained tax and spending policies that boost consumption in ways that reduce the saving rate are likely to lower long-run living standards.” More than two-thirds strongly agreed or agreed. The rest either were uncertain or had no opinion. Remarkably, no one disagreed (IGM Forum, 2013).

10 To be clear, however, the desirability of issuing safe, liquid assets is a justification for gross issuance of debt, not a reason for the government to run deficits. The government could increase the supply of safe, liquid assets and invest the funds in a broad-based market portfolio. That is, saying that the government should issue safe, liquid assets is different from saying that the government should postpone paying for its current programs and instead should raise burdens on future generations.
spending benefit. As a result, national saving—the sum of private and public saving—falls. Once national saving falls, future national income will fall; it is only a question of how. If it generates higher interest rates, government borrowing will crowd out domestic investment, and future output will be lower than it otherwise would have been. Even if interest rates don’t rise at all, future national income still falls. The increase in government borrowing would be financed, in this case, by increased borrowing from abroad. That allows the country to maintain its current investment and output path, but it still causes future income to decline, since a larger share of that output would be diverted to repaying foreign capital holders.11

There is abundant empirical evidence consistent with these views—that sustained deficits and high debt reduce national saving, investment, and growth, and raise capital inflows and interest rates.12 These effects can be substantial. Extrapolating from the empirical and simulation literatures, a reduction of 60 percentage points in the debt-to-GDP ratio (e.g., from 160% to 100%) would raise the real annual growth rate by 0.6 to 1.2 percentage points according to a study by IMF researchers (Woo & Kumar, 2015). It would raise the long-term GNP level by estimates that range from 4.0% (CBO, 2016a) to 5.7% (Elmendorf & Mankiw, 1999), and 4.2% to 10.5% according to another CBO study (Page & Santoro, 2010), depending on how other policies change.

Likewise, in my recent book, Fiscal Therapy, I propose a series of policy changes that would reduce the debt-to-GDP ratio to 60% by 2050, compared to a 180% figure that would be reached under the assumptions in the analysis (Gale, 2019). Analysts at the Penn-Wharton Budget Model (PWBM) estimate that these changes would raise GDP by 7% and GNP by 8% (Ricco, Prisinzano, & Shin, 2019). This implies that a reduction in the debt-to-GDP ratio of 60 percentage points, coupled with similar policies, would raise long-term GDP by 3.5% and GNP by 4%.

All these estimates are based on extrapolating the effects of large changes in debt from evidence on the effects of smaller changes in debt. Thus, the usual cautions about out-of-sample predictions apply. If anything, though, the extrapolations are likely to understate the effects of large debt changes.13

---

11 GDP measures the output produced in the country. Gross national product (GNP) measures the income that accrues to Americans. GNP equals GDP less the income earned in the United States by foreigners plus the foreign income earned by Americans.

12 For the impact of debt on growth and investment, see Baum, Checherita-Westphal, and Rother (2012); Caner, Grennes, and Koehler-Geib (2010); Cecchetti, Mohanty, and Zampolli (2011); Chudik, Mohaddes, Pesaran, and Raissi (2015); Wilson et al. (2012); and Woo and Kumar (2015). For debt and interest rates, see Engen and Hubbard (2005); Gale and Orszag (2004); Gamber and Seliski (2019); Krishnamurthy and Vissing-Jorgensen (2012); Laubach (2009); and Tedeschi (2019). For debt and capital flows, see Chinn and Ito (2005, 2008); Chinn et al. (2011); and Huntley (2014).

13 The data suggest that while low levels of debt do not necessarily hamper economic performance and may even bolster the economy, high debt tends to reduce economic growth, suggesting the marginal effect of added debt on the economy may be nonlinear, and it may be rising with higher levels of debt. See Baum, Checherita-Westphal, and Rother (2013); Caner, Grennes, and Koehler-Geib (2010); Cecchetti, Mohanty, and Zampolli (2011); Chudik et al. (2015); Reinhart, Reinhart, and Rogoff (2012); Wilson et al. (2012); and Woo and Kumar (2015).
Notably, short-term empirical patterns tell the same story as the long-term simulations. Figure 9, for example, shows that between 1950 and 2018, annual federal saving (that is, the opposite of the federal deficit) correlates closely with national saving and national investment. *Controlling for the business cycle* (by including the unemployment rate), raising deficits by 1% of the economy reduces both national saving and national investment by about 1% of the economy.\(^\text{14}\) Note that \(r < g\) for much of this period (Blanchard 2019a, 2019b).

**Figure 9: National Saving, Federal Saving, and Net Domestic Investment, 1950-2018**

![Figure 9: National Saving, Federal Saving, and Net Domestic Investment, 1950-2018](source: Author’s calculation)

Blanchard (2019a, 2019b) emphasizes that, although sustained deficits would crowd out investment and reduce future national *income*, increased deficits would bring about higher *consumption and welfare* for all generations as long as \(r < g\).\(^\text{15}\) That implies that we can borrow and consume more if interest rates stay low forever. But if we accumulate a lot of debt and then rates rise, we will face added burdens. This is the “Deficit Gamble” that Ball, Elmendorf, and Mankiw (1998) describe.

\(^{14}\) This statement is based on linear regressions using annual data on GDP and net national product (NNP) from 1950 to 2018 (Bureau of Economic Analysis, 2019), controlling for the unemployment rate. The impact of federal saving/NNP on national saving/NNP is 1.37 and the impact on investment/NNP is 0.89. All the effects are highly statistically significant. Similar findings hold using GDP instead of NNP.

\(^{15}\) Blanchard (2019a, 2019b) distinguishes two interest rates—on government debt and on risky private capital. In his model, for increased debt unambiguously to make all generations better off, both rates must be below the growth rate.
Of course, it is impossible to know the future path of interest rates with certainty. But Blanchard (2019b) argues that the situation will remain “manageable” as long as interest rates do not rise much above the growth rate. I would note that the interpretation of “manageable” is subjective. As shown above, even if $r < g$ for the next 30 years under the “standard” interest rate projection, net interest payments rise to more than 6% of GDP by 2049. It would be reasonable to conjecture that many people would find that situation problematic. If the interest rate were to rise above the growth rate, interest payments would be even higher.

### 3.2 Debt and Financial Crisis

In recent decades, prominent economists and leading Wall Street figures of both political parties have expressed concern that America could experience a kind of “hard landing” or crisis, similar to what happened in Greece. Nevertheless, I doubt that we’ll see a sudden scenario in the United States in the foreseeable future, for several reasons. Current low interest rates indicate that markets are absorbing recent increases in government debt without fear of future capital flight or default. We undoubtedly have the resources to pay our debt for decades to come. We issue bonds in our own currency (as do Britain and Japan), giving us an important lever of control over our debt, and the dollar is the world’s reserve currency. The United States remains the world’s safest place to invest; even after the financial crisis that began here in 2007 and spread across the world, investors flooded U.S. markets in search of safe assets, helping to keep interest rates low.

To be sure, policymakers could create an emergency by forcing a default on the country’s debt, as right-wing leaders and commentators threatened to bring about during the debt ceiling standoffs in 2011 and 2013 (Bartlett, 2013; Weisman, 2013). An intentional default would be a big mistake. A financial crisis would turn out poorly, of course, and it would make the need to address the fiscal challenge even more compelling.

But I believe that focusing on the potential for a crisis is misleading, in two ways. First, it seems like an extremely remote possibility. Second, it implicitly suggests that the potential to cause a crisis is the reason we should care about debt. In contrast, the key point in my view is that even if a crisis does not materialize, the United States still faces a debt problem. It’s just one that’s growing gradually. This may be less exciting than a crisis, but it can still be plenty damaging.

---

16 If a crisis were to arise, it could spread quickly. Global financial markets can respond to events virtually instantaneously, and policymakers can lose control of things just as quickly. As Rudiger Dornbusch (1997) said of Mexico’s financial crisis of the 1990s, “The crisis takes a much longer time coming than you think, and then it happens much faster than you would have thought… It took forever and then it took a night.”

17 The rise of cryptocurrencies and efforts to avoid U.S. trade sanctions may be creating threats to the use of the dollar as a reserve currency (Michaels, 2019; Scheck & Hope, 2019).
3.3 Intergenerational Burdens

Besides its impact on overall macroeconomic performance, issuing debt shifts the burden of financing government to future generations (assuming the increase in debt is financing government consumption or transfer payments that will largely raise private consumption as opposed to investments that will pay dividends in the future). There is a natural tendency to think that future generations will be better off than we are, and therefore that pushing the debt forward would simply be asking more affluent people than ourselves to bear the burden. It is not clear, however, how much better off future generations will be relative to current generations. Absolute income mobility has been declining over the last few decades (Figure 10). Among people born in 1940, more than 90% had higher real income as (young) adults than their parents did.

Among those born in 1980, only slightly more than half had higher real income than their parents (Chetty et al., 2017). If that pattern continues, many members of future generations will be worse off in absolute terms than their parents and thus will be less well-prepared to address a higher debt burden than earlier generations.

![Figure 10: Percent of Children Earning More than Their Parents](image)

**Source:** Author’s calculation
3.4 Debt and Politics

High and rising debt can affect political choices as well. In the face of fiscal pressures, policymakers will naturally be less willing to raise debt or deficits further and perhaps more willing to impose PAYGO requirements on new programs (Romer & Romer, 2017). This will make it harder to enact new initiatives that respond to economic, social, military, or other needs. For example, countries with low debt-to-GDP ratios at the beginning of a financial crisis tend to have smaller declines in output than countries with higher debt loads. The reason is that countries with low debt-to-GDP ratios are more willing to enact expansionary policies (Romer & Romer, 2019).18

Another dimension of the impact of debt on political choices occurs through the effects of rising interest payments, which will require either higher taxes, cuts in other spending, or acceptance of higher deficits. This concern is particularly salient under the standard interest rate scenario, where interest payments rise from 1.8% of GDP in 2019 to 6.2% in 2049. By 2049, even with r < g throughout the entire projection period, interest payments would exceed the sum of all discretionary spending. The explosive growth in interest payments, assuming no policy changes, highlights the importance of getting the debt under control sooner rather than later.

More broadly, high debt may reduce America’s global standing in political and military terms. The precise mechanism through which this might occur is unclear, but the general idea is that economic strength and political strength go hand in hand.19

4. Fiscal Targets and The Fiscal Gap

4.1 Alternative Targets

Determining optimal fiscal policy is a task fraught with uncertainty. While it is hard to argue that the current debt path is optimal, it is even more difficult to ascertain what would be optimal.

The long-term solution should be a stable situation—defined either in terms of debt or interest payments, relative to GDP. But economic theory offers such a wide range of plausible estimates for the optimal debt-to-GDP ratio that it’s difficult to...
reach conclusions based on first principles (Aiyagari & McGrattan, 1997; Blanchard 2019a; Peterman & Sager, 2017). Empirically, there are three natural benchmarks to consider. The current ratio is about 78%. The maximum historical ratio was 106%, during World War II. The average ratio from the 50 years between 1957 and 2007 was 36%. This period, which might be described as “normal,” begins after debt as a share of the economy was cut in half relative to World War II levels and ends before the financial crisis.

Both objective and subjective factors influence the determination of the optimal level. Objectively, the more that debt hurts long-term growth, the lower the optimal level. A key parameter is the relation between interest rates and growth rates. Higher interest rates raise the cost of financing and reduce the optimal debt level. Faster economic growth—through productivity growth or labor force expansion—raises the optimal debt-to-GDP ratio, just like a family that expects its future income to rise can responsibly assume more debt.

If \( r < g \), the government can rollover existing debt without raising the debt-to-GDP ratio (Blanchard, 2019a). That, by itself, does not solve the fiscal problem, however. Because U.S. projected primary deficits are positive and large (Figure 2), the U.S. debt-to-GDP ratio is projected to rise continually, as noted above, even with \( r < g \).

Blanchard (2019a) highlights conditions—essentially, that the relevant interest rate be below the growth rate—under which a sustained increase in deficits and debt could make all generations better off. If those conditions do not hold, however—and it is impossible to be certain they will hold in the future—the key subjective issue is how much of the debt burden each generation should bear. Generally, the costs of debt reduction come before the benefits fully kick in, since the benefits, namely higher economic growth, accumulate slowly over time. As a result, society’s willingness to assume current costs for future gains will affect the optimal choice of debt. Deciding which individuals within each generation should bear the burden of debt is a related question that serves to link debates about trends in income inequality and economic opportunity with debates about fiscal consolidation.

The arguments for a higher optimal debt-to-GDP ratio in the future compared to various points in the past focus on two points: The interest rate on government debt is projected to be less than the economic growth throughout the projection period, and the projected baseline debt levels are already so high, making transitions to lower debt levels more expensive. The arguments against letting the ratio rise too high include the effects described in the previous section.

As an alternative metric to the debt-to-GDP ratio, it is not unreasonable to focus on interest payments as a share of GDP. Interest payments, as mentioned above, were 1.6% of GDP in 2018, averaged 1.8% of GDP since 1947, and peaked in 1991 at about 3.2% of GDP. One caveat, however, is that using an interest-payment-to-GDP...
target could require abrupt changes in the budget. For example, a change in the interest rate would have a much larger impact on interest payments as a share of the economy than on the debt-to-GDP ratio.\textsuperscript{20}

4.2 Fiscal Gap Estimates

The “fiscal gap” measures how much policy would have to change on net for the government to reach a given fiscal target (debt-to-GDP or interest payments-to-GDP) by a particular year, given the date when the initiatives are first implemented.\textsuperscript{21} For example, with standard interest rates, achieving a debt target of 60% of GDP by 2049 would require a combination of permanent tax increases and spending cuts that equal 3.8% of GDP per year if the changes start in 2021 (Table 1). This would equal about $800 billion per year in 2019 dollars, with the dollar figure rising at the same rate as GDP in future years. There are many ways to make those changes, but they all involve enormous changes in policy. In 2019, for example, $800 billion represents a 46% increase in income tax revenues, a 23% increase in all federal taxes, or a 20% cut in all non-interest federal spending.

Increasing the debt target reduces the fiscal gap. With a 2049 debt target of 100% of GDP and standard interest rates, the fiscal gap is 2.3% of GDP; if the target is 140% of GDP, the fiscal gap is 0.8% of GDP.

Delaying action generally increases the fiscal gap (even though $r < g$). With standard interest rates, the fiscal gap is 2.7% of GDP for a debt goal of 100% of GDP if policy changes do not begin until 2025, or 3.2% if changes begin in 2029, compared to 2.3% if they start in 2021.

As noted, the target could be in terms of interest payments instead of debt. To keep interest payments in 2049 no higher than their historical peak of 3.2% of GDP would require policy adjustments equal to 2.8% of GDP if policy changes begin in 2021.

The interest rate assumption has a huge impact on the fiscal estimates. With flat interest rates, the fiscal gap falls to 3.0% of GDP with a 60% target (compared to 3.8% under standard rates) and 1.3% of GDP with a 100% target (compared to 2.3% under standard rates). If action is delayed until 2029, the fiscal gap rises, but only slightly, to 1.6% of GDP.

\textsuperscript{20} For example, consider an economy in steady state with debt/GDP = 100%, $r = .02$, $g = .04$, and the primary deficit = 2% of GDP. Interest payments would equal 2% of GDP. The deficit (interest payments plus the primary deficit) would equal 4% of GDP. If $r$ rose to .03, interest payments would rise to 3% of GDP. Stabilizing debt/GDP at its previous value of 100% would require reducing the primary deficit by 1% of GDP via tax increases or spending cuts. In contrast, stabilizing interest payments at the previous value of 2% of GDP would require cutting the debt-to-GDP ratio to 66.7%—that is, it would require raising taxes or cutting spending by 33.3% of GDP. Even if the interest rate target were phased in over several years, the required adjustments would be substantially larger than those required by maintaining a debt-to-GDP target.

\textsuperscript{21} The fiscal gap methodology was developed by Auerbach (1994) and has been used extensively. For a recent example, see Auerbach, Gale, and Krupkin (2018, 2019).
With a 140 percent-of-GDP debt target or a 3.2 percent-of-GDP interest payment target, the fiscal gap is negative—that is, policymakers could expand deficits under this scenario and still reach the target for debt or interest payments.

Long-term budget projections are, of course, uncertain. The CBO (2019c) provides a sense of the range of the uncertainty, noting that if interest rates turn out to be 1 percentage point higher (lower) than currently projected, while still using their baseline assumptions for primary spending and revenues, the debt-to-GDP ratio in 2049 would be 55 percentage points higher (37 percentage points lower). Likewise, if total factor productivity growth were 0.5 percentage points higher (lower) than currently projected, the debt-to-GDP ratio would be 38 percentage points lower (41 percentage points higher). To be clear, these are very large deviations in values for interest rates and productivity, relative to their projected values, so the sensitivity analyses suggests that rising future deficits are extremely likely to occur.

5. Fiscal Policy in the Short- and Long-Run

5.1 Short-Term Policy

Current deficits are certainly not optimal in any first-best sense. Full-employment deficits of 5% of GDP are, and should be, rare. As John Maynard Keynes said, “the boom, not the slump, is the right time for austerity at Treasury.” Yet, perhaps surprisingly, the case for adjusting short-term deficits from their current path is weak, unless there is a recession, in which case short-term expansionary policy is appropriate.

The case for cutting short-term deficits is almost nil, in my view. Admittedly, cutting the current deficit would reduce future debt accumulation, holding the economy constant. But with the economy perhaps becoming fragile and interest rates already low, there is not much room for monetary policy to respond to the reduction in aggregate demand that would come from reducing current deficits. In any case, the current deficit is not the problem; the long-term path is.

The case for raising short-term deficits to finance new investments is based on the ideas that we need new investment in infrastructure, research and development, and social policy initiatives that generate human capital; the macroeconomy may need a boost in the near future; and lower interest rates make more government investment projects beneficial and make deficit financing more attractive.\(^{22}\)

22 There are issues regarding what constitutes an investment, and whether that includes only traditional items like infrastructure, research and development, and human capital, or if it is extended to include programs that provide nutrition, child support, job training, etc. I favor the broader definition on economic grounds but recognize the political complications that arise when investment is defined broadly.
There are some natural caveats to expanding deficit-financed investment, however. First, the government still needs to be able to identify and implement high-value investments. Low interest rates do not justify “bridges to nowhere.” The social opportunity cost of the funds is related to the market return on capital, not the government borrowing rate, and it should account for the irreversibility of investments (Auerbach, 2019). Second, both the CBO and the PWBM estimate that a tax-funded infrastructure program would boost the economy more than a deficit-financed program (Congressional Budget Office, 2016b; Penn-Wharton Budget Model, 2018). Thus, I support a stronger investment program right now and would prefer that it be funded; however, consistent with the ideas that the short-term deficit is not the problem and that the nation needs new investments, I would support a well-designed investment program that is deficit-financed in the short-term.

Like most other people, I do not advocate for deficit-financed increases in non-investment spending currently, even with $r < g$. In the absence of a recession, almost no one seems to think that we should increase deficit-financed spending on non-investment items, even among those who emphasize the role of low interest rates. Krugman (2019) notes that “You don’t have to be a deficit scold or debt-worrier to believe that really big progressive programs will require major new revenue sources.” Furman and Summers (2019) argue for PAYGO to apply to non-investment spending. Sarin and Summers (2019a, 2019b) argue for substantial tax increases on the wealthy, presumably to finance new spending, not to reduce deficits.

At one level, this consensus is not surprising, given the size of the current and projected deficits. On the other hand, it is puzzling given the emphasis placed on low interest rates. After all, the conditions in Blanchard (2019a)—and earlier in Diamond (1965)—under which low interest rates imply that higher deficits are optimal for current and future generations, assume that all government spending is non-investment.

A related question is whether Congress should impose PAYGO rules, which require that new tax cuts or spending increases be “paid for” with other policy changes. The argument in favor is straightforward: As Furman and Summers (2019) put it, if you find yourself in a hole, you should stop digging. And, it is always appropriate to make it salient for political leaders that programs eventually must be paid for.

But there are arguments against PAYGO rules as well. First, they make it very hard to do anything new, but they give existing programs a free pass on accountability. The argument in favor of PAYGO is often stated as a variant of “anything worth doing is worth paying for.” If that is true, it is not clear why it should not apply to existing programs. Second, the long-term projections above are very close to being consistent with PAYGO; that is, PAYGO itself does very little to alter the long-term path from the projections. Third, the political parties may be more willing to negotiate a long-term fiscal agreement if PAYGO did not exist, that is, if both sides could act (or could credibly threaten to act) recklessly with regard to short-term spending and tax cuts.
5.2 Long-Term Policy

There is widespread agreement that the long-term budget outlook is unsustainable—even if interest rates stay below the economic growth rate—and needs to be addressed at some point. The debate is whether we should be doing anything about it now. To be clear, “doing anything” means enacting a set of changes in the near term that are implemented over the medium- and long-term. It does not mean cutting debt immediately.

Clearly, the future fiscal situation depends significantly on interest rates—and of course on other factors including the state of the economy and policy makers’ choices. The question is, essentially, whether we should buy any partial insurance now via a phased-in debt reduction package against potentially adverse future fiscal outcomes.

Elmendorf (2019) acknowledges that the long-term fiscal outlook is unsustainable and must be addressed at some point, but argues against enacting a gradual, long-term budget package in the near term. With lower interest rates, he notes, the nation can carry more debt than previously thought. Moreover, he is concerned that implementing a gradual, long-term debt reduction package would hurt the short-term economy by reducing current aggregate demand and interest rates, which would both make a recession more likely and make it harder for monetary authorities to respond to a recession.23 Similarly, Furman and Summers (2019, p.94) argue that “if the debt becomes a problem, interest rates will rise … but even if that happens, it is not likely to cost so much that it would be worth paying a definite cost today, to prevent the small chance of a problem in the future.” In short, there is a long-term problem, but these authors argue it is not worth trying to fix it currently. It would be better to let the problem ride.

This could be called the “St. Augustine” approach: “Give me chastity and continency, only not yet.” (Pusey, 1909-14; Tax Policy Center, 2019). If so, the St. Augustine view begs for a “Hillel” response: “If not now, when?” (Hillel). After all, the economy has been strong in recent years and full employment deficits are high. As President Kennedy said, “the time to fix the roof is when the sun is shining.” If the answer is “when interest rates rise” (as Furman and Summers, 2019 note above), this is a slippery slope. Certainly, if \( r \) rose to levels above \( g \), some sort of long-term, fiscal containment would be clearly needed, because the debt-to-GDP ratio would rise even more rapidly than shown above. But if \( r \) remains below \( g \) and net interest payments rise to exceed 6% of GDP, as in the standard interest rate scenario, when should deficit reduction begin?

---

23 Elmendorf (2019) also argues that in the current political environment, most of the fiscal adjustment would come on the spending side, when he believes it would more appropriate to have more of adjustment come on the revenue side, and that focusing on deficit reduction would take policymakers’ attention away from other key issues.
An even stronger view could be extracted from Rachel and Summers (2019), who present evidence that rising public debt has helped the economy in the recent past and is boosting the economy now. I have not yet seen anyone argue that the current projected long-term debt path will be necessary to bolster an economy suffering from long-term, secular stagnation, but that seems to be a possible implication of their findings.

My own views align with a different statement by Furman and Summers (2019, p. 91): “The optimal policy from an economic standpoint would be to gradually phase in spending cuts or tax increases at a rate that would prevent perpetual growth in the national debt as a share of the economy but that would avoid doing serious harm to economic demand along the way.” Although they immediately dismiss their idea as too “nuanced” for the political system, that does not mean the idea is wrong, and I do not see why it is too complex. Social Security reform in 1983, for example, phased in a variety of gradual changes, including to the full retirement age, a change that is still being implemented.

The main cost of enacting a gradual, long-term debt reduction plan seems to be that it could hurt the current economy if the changes took place too quickly. In principle, that seems like an easy problem to address, by phasing in the changes slowly. Of course, when policymakers enacted deficit reductions in 2011 and 2013 in the midst of weak economies, they imposed the changes immediately, which was a mistake. The “gradual” part of a debt-reduction plan matters.

The benefits of having a plan in place seem clear. First, it would be a first step toward dealing with a long-term problem that everyone acknowledges exists. As Furman and Summers (2019, p.90) note, “since economists aren’t sure just how costly large deficits are, it would be prudent to keep government debt in check in case they turn out to be more harmful than expected.” Second, it provides a form of insurance, should interest rates begin to rise. There is no guarantee that r will stay low, especially given the projected increase in debt. Given the already high and rising level of debt relative to GDP, the federal budget has rarely been more vulnerable to interest rate shocks. Third, it would provide policymakers with some assurance about the fiscal path and thus allow them to address new problems or issues from a framework that is fiscally stable. Fourth, it would reduce fiscal policy uncertainty and provide time for businesses, investors, and taxpayers to adjust their behavior to new rules. The last point is particularly important because abrupt changes are likely to generate backlash and thus may be more likely to be undone by future policy actions.

Another way to “buy insurance” against interest rate fluctuations would be to lengthen the maturity structure of the federal debt. Currently, about 30% of privately held public debt will mature within a year, 70% within 5 years, and 90% within 10 years (“Table FD-5,” 2019). Issuing more bonds with longer maturities as the size of the debt rises makes
sense conceptually and matches the historical pattern. Doing so can help reduce the sensitivity of the budget to short-term interest rate risk. In the past, the argument against lengthening the maturity was that it would raise interest payments, given long-term interest rates are typically higher than short-term rates. Currently, however, the yield curve is relatively flat, which makes lengthening the average maturity of federal debt more palatable (Figure 11). But it is also important to consider the implications of changing the maturity on the overall financial system and on the Fed’s ability to conduct quantitative easing, should the economy turn down again.24

Figure 11: Yield Curve, Closest to May 31 or June 1, Selected Years 1990-2019

![Yield Curve Graph](image)

Source: Author’s calculation

6. Conclusion

Policymakers face a combination of high and rising debt but relatively low interest rates. Low rates help mitigate the costs of debt, but the long-term fiscal outlook is troublesome even if interest rates stay below the growth rate for the next 30 years. In a similar situation after World War II, the United States ran extremely small primary deficits on average for 3 decades. In contrast, future primary deficits are projected to be both sustained and persistent, and interest payments are projected to rise inexorably. To address the fiscal imbalance, policymakers should enact now a gradual,

---

24 For a more complete discussion of these issues, see the excellent contributions in Wessel (2015).
phased-in, long-term plan that would reduce primary deficits substantially over time and eventually stabilize the debt-to-GDP ratio at a plausible level.

**2025 Table 1. Fiscal Gaps and Net Interest Through 2049 With Various Debt Targets and Interest Rates, Starting in 2021 (Percent of GDP)**

<table>
<thead>
<tr>
<th></th>
<th>Fiscal Gaps</th>
<th>Net Interest in 2049 Under Target</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard Interest Rates</td>
<td>Flat Interest Rates</td>
</tr>
<tr>
<td>Debt target = 60%</td>
<td>3.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Debt target = 100%</td>
<td>2.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Start in 2025</td>
<td>2.7</td>
<td>1.4</td>
</tr>
<tr>
<td>Start in 2029</td>
<td>3.2</td>
<td>1.6</td>
</tr>
<tr>
<td>Debt target = 140%</td>
<td>0.8</td>
<td>-0.5</td>
</tr>
<tr>
<td>Net Interest = 3.2% of GDP in 2049</td>
<td>2.8</td>
<td>-0.4</td>
</tr>
</tbody>
</table>

Memorandum:
- Baseline 2049 debt: 169, 134
- Baseline 2049 net interest: 6.2, 3.1
- Debt goal associated with 3.2% of GDP net interest: 88, 138
Appendix A

Constructing a “Current Policy” Budget Baseline

Constructing a budget projection is part art and part science. The Congressional Budget Offices “current law” projections essentially assume that Congress does (almost) nothing in the future. For example, the projections assume that temporary tax changes expire as scheduled, mandatory programs are reauthorized as scheduled, and discretionary spending follows the caps set forth in the Budget Control Act of 2011 (which were modified in subsequent legislation) through 2021 and remains constant in real terms thereafter (CBO, 2019a, 2019b).

In contrast, I construct a “current policy” baseline that shows where the budget is headed if we stay on what, in my judgment, is our current path (Appendix Table 1). This is essentially what would happen if Congress follows a “business as usual” approach. My projections start with the CBO’s current law estimates and make a series of adjustments. These adjustments are not policy recommendations; they simply show the effects of what I view as a continuation of current policies. In many cases, I utilize estimates that CBO itself provides of alternative policy options.

I assume that, as it has done in the past, Congress makes major temporary tax-cut provisions permanent, including the temporary provisions in the 2017 tax act. I also assume that enacted tax provisions for which implementation has already been delayed will be permanently delayed (i.e., the provisions will be cancelled and never take effect). This includes the medical device excise tax and the tax on high-premium insurance (the “Cadillac Tax”) that were enacted as part of the Affordable Care Act. With bipartisan support, the implementation of these taxes was postponed by two years in the Protecting Americans from Tax Hikes Act in December 2015 and by another two years in the Extension of Continuing Appropriations Act of 2018.

On the spending side, as mentioned above, the CBO sets discretionary spending through 2021 at the levels created by the discretionary spending caps and sequestration procedures (as imposed in the Budget Control Act of 2011 and modified by the Bipartisan Budget Acts of 2013, 2015, and 2018) and then allows...

---

25 The “current law” scenario does assume that the debt limit is raised as needed.

26 Examples of major expiring provisions in the 2017 tax act include “100% bonus depreciation” (expensing of business investment in qualifying equipment), the marginal individual rate cuts, the increased standard deduction, the repeal of personal exemptions, the increased estate tax exemption, the cap on state and local tax deductions, and the 20% deduction for certain pass-through income. Examples of expiring provisions outside of the 2017 tax act include tax credits for biodiesel and alternative fuel mixtures and the deduction for mortgage insurance premiums.

27 The revenue adjustments also affect refundable tax credits, which, in accordance with the CBO (2019a, 2019b), is considered an effect on outlays.
them rise with inflation. I allow defense spending to rise with inflation, starting in 2020, so that real defense expenditures remain constant at 2019 levels. I allow non-defense discretionary spending to rise with the rate of inflation and the rate of population growth, so that real, per-capita spending remains constant at its 2019 level. Both assumptions are meant to reflect a rough approximation of a budget that maintains current services. For defense, largely a non-rival public good, it seems reasonable to assume that current services can be maintained without regard to population over the short-term. For non-defense programs, it is more likely that maintaining current services requires a population adjustment.

In aggregate, my ten-year current policy baseline follows the CBO’s (2019b) alternative fiscal scenario, except for the population adjustment I make for non-defense discretionary spending.

The CBO (2019b) explains that the deficit for fiscal year 2029 will be about $93 billion lower than would otherwise be expected because October 1, 2028 (the beginning of fiscal year 2029) will fall on a weekend, thus pushing some October payments (mostly for Medicare) up to the end of September in the previous fiscal year. As a result, the deficit in 2028 will be larger than otherwise expected. Of these $93 billion in payments, $64 billion applies to Medicare. Similar adjustments affect spending in fiscal years 2022-2024. Figures in this paper display the adjusted baselines that exclude the effects of these timing shifts.

Looking only at the next ten years gives an incomplete and overly optimistic picture of the fiscal outlook, even with adjustments made to characterize current policy. After the initial 10 years, I use long-term economic growth assumptions implied in CBO (2019c) without macroeconomic feedback. Over the 2030-2049 period, the average nominal economic growth rate is about 4%.

For Medicare and OASDI, I project all elements of spending and dedicated revenues (payroll taxes, income taxes on benefits, premiums and contributions from states) using the growth rates in the intermediate projections in the 2019 Trustees Reports for the period between 2030 and 2049 (The Board of Trustees, 2019; The Boards of Trustees, 2019). To account for the timing shifts discussed above, Medicare spending in 2030 is based on the growth rate of spending between 2029 and 2030 according to Boards of Trustees (2019) applied to the adjusted Medicare estimate for 2029 from the CBO (2019b). For Medicaid, CHIP, and exchange subsidies, I use growth rates implied by the CBO’s most recent, long-term static projections (CBO 2019c) through 2049.

---

28 The CBO (2019a, 2019b) uses a mix of the employment cost index and the GDP price index to measure inflation.

29 For simplicity, I use the same nominal adjustment for the timing shifts in each scenario.

30 Details of the computations are available from the authors upon request.

31 The static projections are based on macroeconomic forecasts for a constant debt-to-GDP ratio and constant marginal tax rates after 2029, that is, excluding the negative effects of economic policy during this period.
As in the first 10 years, I hold non-defense discretionary spending constant in real, per capita terms and defense spending constant in real terms after 2029. I also hold mandatory spending for programs other than Social Security, Medicare, Medicaid, CHIP, and exchange subsidies constant in real per capita terms.

On the revenue side, I allow income taxes other than those outlined above on Social Security and Medicare benefits to grow with “bracket creep” according to the CBO (2019c). I assume that all other revenues (corporate taxes, excise taxes, etc.) remain constant at their 2029 shares of GDP.

I examine two alternative interest rate paths. In the standard approach, I follow the weighted average nominal interest rates on government debt without macroeconomic feedback according to the CBO (2019c) through 2049. Under this path, the weighted average nominal interest rate on government debt rises gradually from 2.4% in 2019, to 3.4% in 2029, to 3.9% in 2049.

In the alternative path (“flat interest rates”), the weighted average nominal interest rate on government debt is constant over the 30-year projection period at its 2019 value of 2.4%.
### Appendix Table 1. Budget Category Assumptions

<table>
<thead>
<tr>
<th>Category</th>
<th>10-YEAR WINDOW</th>
<th>BEYOND</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interest Rate</strong></td>
<td>as implied in CBO 10-year outlook</td>
<td>CBO Long-Term Budget Outlook</td>
</tr>
<tr>
<td><strong>Growth Rate</strong></td>
<td>as reported in CBO 10-year outlook</td>
<td>CBO Long-Term Budget Outlook</td>
</tr>
<tr>
<td><strong>Income Tax</strong></td>
<td>current law + extension of temporary tax provisions and TCJA provisions</td>
<td>current law + extension of temporary tax provisions and TCJA provisions</td>
</tr>
<tr>
<td><strong>Corporate Tax</strong></td>
<td>current law + extension of temporary tax provisions and TCJA provisions</td>
<td>constant share of GDP</td>
</tr>
<tr>
<td><strong>Payroll taxes</strong></td>
<td>current law</td>
<td>grows using assumptions in the Social Security Trustees report</td>
</tr>
<tr>
<td><strong>Other taxes</strong></td>
<td>current law + repeal of certain healthcare taxes</td>
<td>constant share of GDP</td>
</tr>
<tr>
<td><strong>OASDI benefits</strong></td>
<td>current law</td>
<td>grows using assumptions in the Social Security Trustees report</td>
</tr>
<tr>
<td><strong>Medicare</strong></td>
<td>current law</td>
<td>grows using assumptions in the Medicare Trustees report</td>
</tr>
<tr>
<td><strong>Medicaid, CHIP, and Exchange Subsidies</strong></td>
<td>current law</td>
<td>grows using assumptions in CBO Long-Term Budget Outlook</td>
</tr>
<tr>
<td><strong>Other Mandatory</strong></td>
<td>current law</td>
<td>grows with inflation and population</td>
</tr>
<tr>
<td><strong>Defense</strong></td>
<td>grows with inflation</td>
<td>grows with inflation</td>
</tr>
<tr>
<td><strong>Non-Defense Discretionary</strong></td>
<td>grows with inflation and population</td>
<td>grows with inflation and population</td>
</tr>
<tr>
<td><strong>Net Interest</strong></td>
<td>as reported in CBO 10-year outlook + adjustments from above policy changes</td>
<td>calculated from debt and interest rate</td>
</tr>
</tbody>
</table>
Appendix B
Dispelling Some Canards about Debt and the Economy

Despite a broad consensus that the long-term fiscal path described in Section 2 is worrisome or inappropriate, even with $r < g$, there are a few false or misleading claims that merit specific responses.

1. We owe it to ourselves, so it is not a problem.

The first false claim is that public debt is not a problem because “we owe it to ourselves.” By this statement, people mean that public debt is money that one generation borrows and owes to another. How, they ask, can the nation become poorer by owing money to ourselves?

The answer is that the historical evidence discussed above on how deficits affect growth, saving, investment, and interest rates refers—at least in the U.S. case—to debt that we, indeed, largely owed to ourselves. Those deficits and debt affected economic performance, through the channels described above. In addition, future generations will have to finance that debt via higher taxes or lower spending, and those steps will cause pain, especially if we design the policies poorly.

Also, we increasingly do not owe it to ourselves; we also owe it to investors around the world. At the end of 2018, foreign investors held 38% of U.S. federal government debt, an amount equal to 30% of our annual GDP (“The depth and breadth,” 2017). Those figures are substantially higher than in 1980, when foreign investors held about 18% of U.S. public debt.

2. We issue debt in our own currency, so there is not a problem.

Another argument suggesting that our fiscal situation is not a problem is that we print and borrow in our own currency and so can never be forced to default. For instance, in 2011 Warren Buffet said, “the United States is not going to have a debt crisis as long as we keep issuing our debts in our own currency” (Wood, 2011). However, this does not mean that a fiscal problem cannot happen; in 1976, the government of the United Kingdom, which issues its own currency and borrows in its own currency, was forced to borrow $3.9 billion from the International Monetary Fund when the pound rapidly fell in value (“Sterling devalued,” 2019).

The key issue is the costs and benefits of additional debt accumulation, not the limits of federal borrowing. In a recent University of Chicago survey of prominent economists, not one agreed that a country that issues debt in its own currency does not have to worry about deficits (IGM Forum, 2019). In post-survey comments, even adherents to modern, monetary theory, believed that a government that printed its own currency needed to be concerned with its level of debt (Mitchell, 2019).
3. Ricardian Equivalence says that our rising debt profile does not matter.

There is a school of thought that says that deficits do not reduce growth under certain conditions. In particular, the theory behind so-called “Ricardian Equivalence” is that a deficit that is created by a temporary, lump-sum tax cut today and then followed by a temporary, lump-sum tax increase in the future will not have any impact on national saving, investment, growth, or interest rates. The reason is that taxpayers will anticipate that their future tax liabilities will rise by the exact amount (in present value) of the tax cut they receive, and so they will save the entire tax cut in order to pay the future tax increase. Thus, the reduction in government saving due to the tax cut would be exactly offset by the increase in private saving, and there would be no change in national saving. Ricardian Equivalence is named after the nineteenth-century British economist David Ricardo, who did not actually believe in the idea but raised it as a conceptual possibility. It was revived intellectually by Harvard economist Robert Barro (1974) in a famous (in academia, at least) article. While the theory is intellectually elegant, there is significant evidence against it (Bernheim, 1989; Elmendorf & Mankiw, 1999), and, in any case, it does not apply to the situation facing the country—namely, rising long-term deficits and debt-to-GDP ratios that finance government and private consumption. Virtually all economists, including Barro (2012), agree that if current budget projections play out, they will cause long-term economic harm.
References


Bureau of Economic Analysis. (2019). Table 5.1 Saving and Investment by Sector. Retrieved from https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=3&isuri=1&nipa_table_list=137&categories=survey


Fiscal Policy With High Debt and Low Interest Rates


Can Innovation Policy Restore Inclusive Prosperity in America?

AUTHOR
John Van Reenen

ABSTRACT
Technological innovation is the engine of economic growth, and the key to raising living standards over time. America’s role at the forefront of technological change has traditionally gone hand in hand with its position as the dominant superpower. However, U.S. productivity growth has been lackluster for the past decade, median wage growth has stagnated for almost 40 years, and inequality across people and places has soared. Meanwhile, geopolitical rivals, above all China, are making great strides toward challenging America’s position as the dominant technological power in the global economy. The private sector will not solve these problems by itself. What can be done to boost American innovation? In this memo, I argue that three groups of innovation policies are the most effective way to spur U.S. technological progress and productivity growth: tax credits, direct subsidies, and human capital investments. Combining these tools into a Grand Innovation Challenge program would provide an industrial strategy to promote the dual goals of maintaining America’s technological leadership and promoting inclusive growth.

*Massachusetts Institute of Technology Department of Economics and Sloan Management School: vanreene@mit.edu. The research reported in this publication was supported in part by the Sloan Foundation; Smith Richardson Foundation; National Science Foundation; and Schmitt Sciences.
1. Introduction: Overview of the Challenge

America cemented its place as the world’s economic and technological dynamo after the Second World War. Real Gross Domestic Product (GDP) per capita doubled between 1947 to 1973. Although U.S. productivity growth slowed after the 1970s oil shocks (see Figure 1), the period since the Great Recession of 2008-2009 has been particularly disappointing. Even before this most recent slowdown, however, the outcomes in the labor market have been awful among less-educated individuals. Since 1980, men who have less than a college education have experienced falling real wages (see Figure 2). Median real hourly pay among men fell by 6% between 1979 and 2017. The fruits of growth have not only been harvested more slowly, they have also been very unequally shared.

In the long run, innovation is the only way for an advanced country such as the United States to secure sustainable productivity growth. But what are the most effective policies to stimulate innovation? And how can they be shared more widely? This is the focus of my paper.

Before beginning, I start with the obvious question: Why should taxpayers fund innovation through the government?

Figure 1: U.S. Productivity Growth

Source: Jones (2016)

Note: Shaded areas are NBER recessions
1.1 Innovation Creates Growth

A premise of the argument for government intervention is that innovation is an important driver of aggregate growth. Figure 3 shows research and development (R&D) spending\(^1\) as a fraction of GDP for major industrialized countries. Nations that devote more of their national income to R&D tend to be richer (e.g. Jones, 2016). The United States spends more on R&D than any other country ($495.1 billion), which accounts for roughly 28% of global R&D spending ($1.918 trillion) (National Science Board, 2018).

---

\(^{1}\) R&D is only one measure of innovation inputs and is not of course a perfect measure. It should be complemented with other metrics such as broader inputs to the creation of intellectual property and innovation outputs such as patenting, other IP, direct innovation measures. R&D does have the great advantage of being tracked over a long period of time and across countries in a broadly standard way and also measured directly in dollar terms.
Over time, however, the picture is less rosy. The United States has maintained an R&D-to-GDP ratio of 2.5% to 2.7% since 1981. By contrast, other countries, particularly in Asia (Japan, South Korea, and, most recently and spectacularly, China), have been devoting increasing amounts of national income to R&D. Furthermore, although U.S. R&D intensity has been stable since the mid-1960s, the composition of R&D spending has changed dramatically, as government funding has declined and private-sector funding has increased to fill the void (see Figure 4). Government tends to fund higher risk, basic research that private investors are often reluctant to take on. Therefore, public R&D investment tends to produce higher value, high-spillover inventions over a longer period of time. Despite the decline in government R&D funding, the private sector has also invested less in basic research over time (e.g., Arora, Belenzon, & Patacconi, 2018).

It is difficult to establish whether increased R&D has had a causal impact on economy-wide growth. Perhaps rich countries can lavish money on vanity research projects. Or perhaps there is a third factor, such as rising general education, that increases both GDP and R&D, and thus R&D has no direct effect on growth.
To identify the direction of causality between innovation and growth, academic work has focused on data on industries and firms. There is now a substantial body of evidence indicating that R&D and other measures of innovation (such as quality-adjusted patents) do substantially raise productivity growth. Early work, summarized in Griliches (1998) focused on correlations over time whereas more recent work (e.g., Bloom, Schankerman, & Van Reenen, 2013) uses policy experiments to identify the causal impact.

1.2 But Innovation Can Also Increase Inequality

What is the impact of faster technological change on the labor market? The concern that new technologies will lead to mass unemployment has been with us since Ned Ludd apocryphally broke textile machines in 18th-century England. However, three centuries of technological progress have brought us higher incomes without falling employment rates. If anything, the opposite has been true as women entered the workforce en masse in the latter part of the 20th century.

There is more concern that technical change has biased demand toward more highly skilled workers for at least the last 100 years (Goldin & Katz, 2009). The increase in the relative wages of more educated workers in Figure 2 occurred despite a large
increase in the numbers of workers with a Bachelor’s or higher degrees. Many studies have confirmed skill-biased technical change accounts for much of these trends, more so than other factors such as globalization or institutional changes (see the surveys by Acemoglu & Autor, 2011 and Van Reenen, 2011a). Fundamentally, there is a race between technology and education. Technology increases the demand for highly skilled labor, but if the supply of education keeps up, as it did for most of the 20th century in the United States, then wage inequality need not rise. However, if the increase in education slows down, as it did for cohorts entering the labor market from the late 1970s onwards, the wage difference between more and less educated workers will rise.

This poses a challenge for innovation policy. Increasing the speed of technological change will increase growth, and, by increasing the size of the economic pie, this creates opportunities for all to benefit from, whether rich or poor. However, as the pace of technical progress speeds up, this will tend to benefit the more skilled, increasing inequality. This highlights the need for government to have complementary policies to ensure that the fruits of higher growth are shared equitably. Part of this is through taxes and benefits, but part of this is through ensuring continued increases in high-quality education and training for those from less prosperous families and communities.

1.3 Why Should Governments Promote Innovation?

Just because innovation causes growth does not mean that the government should necessarily support it, as market incentives could suffice. However, it is now well recognized that the market will generally fail to provide enough R&D since the knowledge that is created “spills over” from one firm to another. As one firm creates a new technology, other firms will incorporate learning from the original research without having to pay the full cost of R&D. Ideas are promiscuous; even with a well-designed, intellectual property system, the benefits of new ideas are difficult to fully monetize by the original inventor. Therefore, government investment is needed to ensure overall R&D investment reaches its socially optimal level.

There is a long academic literature documenting the existence of these positive spillovers from innovation (e.g., Bloom et al., 2013). Although firms receive some private returns from their R&D, the literature has consistently estimated that social returns to R&D due to spillovers are much higher than private returns, which justifies government-sponsored innovation policy. In the United States, for example, recent estimates suggest that social returns are about four times as large as private returns (e.g., Lucking, Bloom, & Van Reenen, 2018).

There are many other reasons why the amounts of R&D provided by the private sector will not be efficient (duplicative R&D, risk, financial market frictions, short-termism, business stealing, etc.) but knowledge spillovers are the most important reason.
2. Policy Measures to Address the Innovation Challenge

In Bloom, Van Reenen, and Williams (2019), we examine a wide range of innovation policies. Here, I look at three broad classes of policies—tax incentives, direct grants, and investments in skilled human capital—that have proven to be successful. I also discuss some policies that have proven to be less effective in promoting innovation.

2.1 Tax Incentives for R&D

An obvious approach to stimulating more innovation is through an R&D tax incentive to lower the cost of research. President Ronald Reagan introduced the Research and Experimentation Tax Credit in 1981 and most Organisation for Economic Co-operation and Development (OECD) countries have since followed suit. The policy costs U.S. taxpayers about $11.3 billion annually (National Science Board, 2018). The OECD (2018) reports that 33 of the 42 countries they examined provide some material level of R&D tax support. In France, Portugal, and Chile, which have the most generous incentives, tax incentives reduce the costs of R&D by as much as 40%. In contrast, the United States ranks in the bottom third of the OECD in terms of generosity toward R&D credits.

Do R&D tax credits work? In short, the answer seems to be “yes.” We would expect to observe an increase in R&D when its tax price falls. However, this question is of interest to researchers because expert surveys suggest that R&D is driven by advances in basic science and market demand, rather than any fiscal concerns. There are now a large number of studies examining changes in the rules determining the generosity of tax incentives using a variety of data and methodologies (see Becker, 2015, for a survey). Many early studies used cross-country (Bloom, Griffith, & Van Reenen, 2002) or cross-U.S. states data (Wilson, 2009) to examine the relationship between changes in R&D and changes in tax rules. More recent studies use firm-level data and exploit differences in tax rules across firms before an unexpected policy change takes place. For example, firms below a size threshold may receive a more generous tax treatment, so one can compare firms just below and just above the threshold after (and before) the policy to tease out the real policy effect (Dechezleprêtre, Einiö, Martin, Nguyen, & Van Reenen, 2016). The literature on this topic generally concludes that a 10% fall in the tax price of R&D results in at least a 10% increase in R&D in the long run, and usually much more. This suggests that taxpayers get a big bang for their buck on R&D.

A concern for researchers and policymakers alike is that firms may just relabel existing expenditures as “R&D” in order to take advantage of more generous tax breaks. Chen, Liu, Suárez Serrato, and Xu (2018), for example, found substantial relabeling following a change in Chinese corporate tax rules. A direct way to assess the success of the R&D tax credit is to look at other outcomes such as patenting, productivity, or jobs. Encouragingly, these more direct measures also seem to increase (with a lag) following tax changes.
2.2 Direct Government R&D Grants

A disadvantage of tax credits is that they cannot be targeted at those areas where spillovers may be the greatest. One alternative is for the government to provide direct funding, either to academic researchers, such as through the U.S. National Institutes of Health (NIH), to private firms, such as through the Small Business Innovation Research (SBIR) program, or perform R&D directly in government labs.

Evaluating effectiveness in this context is challenging for at least two reasons. First, public research grants usually (and understandably) attempt to target the most promising researchers, the most promising projects, or the most socially important problems. That type of targeting and concentration of resources means that it is often difficult to construct a counterfactual for researchers, firms, or projects that receive public R&D funds. Second, it is often difficult to appropriately account for the potential crowd out (or crowd in) of private R&D by public R&D. That is, if one dollar of public R&D simply displaces another dollar of private R&D that would have otherwise been invested in the same project, then public R&D could have no real effect on overall R&D spending (much less on productivity growth, patents, or other outcomes).

There are several ways that public R&D influences private firms. First, public R&D funding directed to academics can generate spillovers to private firms. Azoulay, Graff Zivin, Li, and Sampat (2019) exploit quasi-experimental variation in NIH funding across research areas to show that a $10 million increase in NIH funding to academics leads to about 2.7 additional patents filed by private firms. Second, private firms themselves sometimes conduct publicly funded R&D. Moretti, Steinwender, Van Reenen, and Warren (2019) use changes in military R&D spending, which is frequently driven by exogenous political changes, to look at the impact of public subsidies for military R&D. They document that a 10% increase in publicly funded R&D (to private firms) results in a 3% increase in private R&D, suggesting that public R&D crowds in private R&D and raises productivity growth. A third example is Howell (2017), who examines outcomes for SBIR energy R&D grant recipients, using a winner versus losers’ comparison. She estimates that early-stage SBIR grants roughly double the probability that a firm receives subsequent venture capital funding, and that receipt of an SBIR grant has positive impacts on revenue and patenting.

Two other aspects of public R&D support are worth mentioning. First, a substantial share of public R&D subsidies go to universities, which is sensible from a policy perspective as spillovers from basic academic research are likely to be much larger than those from near-market applied research. There is certainly a correlation between areas with strong, science-based universities and private-sector innovation (e.g., Silicon Valley, Route 128, etc.). However, these clusters could arise for many reasons. Andrews (2017) provides the best evidence suggesting the existence of a positive causal effect of universities on innovation outcomes. He analyzes the founding of new colleges in the United States between the mid-19th and mid-20th
centuries, comparing counties where colleges were built with second-choice county locations, and documents a 32% increase in long-run patenting in counties where universities were located.

2.3 Human Capital Supply

The policies described above would increase the demand for R&D workers. However, since R&D workers are in short supply, there is a risk that such demand-side policies would bid up the salaries of these highly skilled workers, without necessarily increasing the volume of R&D. This not only increases inequality, but also is a waste of American taxpayers’ tax dollars. Existing estimates of this effect have not found them to be large (e.g., Bloom et al., 2002), perhaps because of skilled immigration. Nevertheless, such general equilibrium effects are always tough to pin down empirically.

A better, long-run way to increase innovation may be to increase the supply of innovative human capital. This increases the volume of innovation directly as skilled workers are more likely to invent, but also indirectly, by reducing the equilibrium cost of R&D workers.

There are a wide range of policy tools that could be employed to increase human capital. Given the extensive evidence for skill-biased technical change, we would expect these policies to stimulate faster technological diffusion. This is because technology and human capital complement each other. More technology increases the demand for skills; for the same reason, more human capital makes it easier to design and implement new technologies. The most direct policy to expand frontier innovation, however, would be to increase the quantity and quality of inventors. There have been many attempts to increase the number of individuals trained in STEM (Science, Technology, Engineering, and Mathematics). Evaluating the success of such policies is very challenging given the fact these policies tend to be economy-wide, with effects that will play out only in the long run. As noted above, several papers look at the location, expansion, and regulation of universities as key suppliers of STEM workers and track their influence on innovation and growth. The overview in Valero and Van Reenen (2019) suggests universities increase local growth through a variety of mechanisms, including the increase of STEM workers and their subsequent innovation. Other papers using more precise natural experiments also find grounds for optimism that increasing the supply of STEM workers raises innovation (Hausman, 2018; Andrews, 2019; Toivanen & Väänänen, 2015; Bianchi & Giorcelli, 2018).

Another source of innovation-relevant human capital is skilled immigration. Historically, America has had a relatively open immigration policy that has helped to make the nation a magnet for global talent. Immigrants make up only 18% of the labor force aged 25 or more, but constitute 26% of the STEM workforce, own 28% of higher quality patents, and hold 31% of PhDs (Shambaugh, Nunn, & Portman,
Much research supports the idea that immigration boosts innovation. For example, using state panel data from 1940–2000, Hunt and Gauthier-Loiselle (2010) document that a one percentage point increase in the share of college graduates who are immigrants increases patents per capita by 9% to 18%.

Another way to increase the quality of the supply of R&D talent is to consider the barriers that talented people face when becoming inventors in the first place. A growing body of literature matches administrative data on income to an individual inventor’s name on patents and finds that children born in low-income families, women, and minorities face important barriers to becoming successful inventors (“Lost Einsteins”). Bell, Chetty, Jaravel, Petkova, and Van Reenen (2019a), for example, document that American children born into the top 1% of the parental income distribution are 10 times more likely to grow up to be inventors (as measured by being named as an inventor on a patent application or grant) than are those born in the bottom half of the distribution. The majority of this correlation is unrelated to ability and, instead, is causally related to the extent to which a child is exposed to inventors during childhood, such as through their parents, social networks, and neighborhoods. Lack of exposure and role models also seems to be a factor behind the relatively low fraction of women and minorities becoming inventors. These barriers can be reduced through improving school quality in poor neighborhoods and greater exposure to role models and mentors, especially among children who show early signs of STEM skill potential. Bell et al. (2019b) suggest that such policies could quadruple long-run invention rates in the United States.

2.4 Policies That Don’t Increase Innovation

There are large numbers of other policies that have been tried, but failed to significantly promote innovative activity. One example is patent boxes, which are special tax regimes that apply a lower tax rate to revenues linked to patents relative to other commercial revenues. By the end of 2015, patent boxes (or similarly structured intellectual property tax incentives) were used in 16 OECD countries (Guenther, 2017). Although patent box schemes purport to be a way of incentivizing R&D, in practice they induce tax competition by encouraging firms to shift their intellectual property royalties into different tax jurisdictions. In particular, multinational firms have considerable leeway in deciding where they will book their taxable income from intellectual property. Patent boxes provide a system through which they can manipulate stated revenues from patents to minimize their global tax burden (Griffith, Miller, & O’Connell, 2011). Although it may be attractive and effective (see Choi, 2019) for governments to use patent box policies to collect footloose tax revenues, such policies do not have much effect on the real location or the quantity of either

---

research and development or innovation. Gaessler, Hall, and Harhoff (2018) find a small effect of the introduction of patent boxes in several countries in the European Union on transfers of patent ownership, but zero effect on real invention.

In recent years, cuts to the top rates of individual income tax have been suggested as an effective way to incentivize innovation. Bell et al. (2019b) argue that lower top tax rates are unlikely to generate substantially large numbers of new inventors. One reason is that Bell et al. (2019a) documented that exposure to the possibility of becoming an inventor at an early age is an important driving force behind the chances of growing up to be an inventor. Changing top tax rates does not change this. The fact that the Bush top tax cuts did not produce an innovation boom should also give one pause for thought over top-rate tax cuts as an innovation policy. Akcigit, Grigsby, Nicholas, and Stantcheva (2018) argue that lower income taxes across U.S. states raise innovation, but they cannot rule out that this increase may come from the movement of inventors around the United States (see Moretti & Wilson, 2017).

2.5 Summary of Innovation Policies

Today, U.S. federal spending on R&D is about 0.7% of economic output, compared to its peak in 1964 of about 2%. In today’s dollars, the United States spends roughly $240 billion less per year on R&D than it did at its peak. Increasing R&D investment by $100 billion would represent one-half of 1% of GDP and would be transformative for the future of U.S. innovation.

These resources should be spent on the three policy areas identified above, although the timing and rate of return would vary across investments. In the near term, relaxing rules on skilled immigration would have an immediate and near-costless impact. Increasing the generosity of R&D tax credits could also produce quick wins in terms of total, private R&D investment. Directed R&D grants would have a medium-term impact, while human capital investments would have the longest and largest expected return.

3. Are There Lessons From East Asia on Industrial Policy?

3.1 Mission-Oriented Policies

Economists are traditionally skeptical about industrial policy. The conventional view is that markets are generally efficient and even when they are not, governments rarely have the nimbleness and foresight to effectively intervene. In addition, this assumes that bureaucrats are well intentioned and not are captured by vested interests. The experience of European and Latin American industrial policies in which governments
threw money at “national champions” (such as the failed British Leyland in the U.K. auto industry) is not a promising model.

Two things have changed in recent years, however. First, there is more causal evidence on the positive effects of place-based, industrial policies (e.g. Criscuolo, Martin, Overman, & Van Reenen, 2019). Secondly, the slowdown of growth in Western countries and the perceived success of such policies in East Asia has caused some to re-evaluate the case for industrial policy (Rodrik, 2015). China looms large, and its success in science should not be underestimated. For example, Figure 3 showed that in the last decade alone, Chinese R&D grew from 0.5% of GDP in 1996 to 2.1% in 2017. In 1990, China produced only 1.2% of the world’s scientific papers, whereas the United States produced 32%. By 2016, China had surpassed the United States, producing 426,000 compared to our 409,000. The average quality of research papers (as measured by citations) written by Chinese scientists quadrupled over the same period, whilst the quality of those written by American experts declined slightly (Tollefson, 2018).

Drawing on this work, an industrial policy could focus on innovation. There have been many such “mission-oriented” policies in the United States around defense (e.g. the Defense Advanced Research Projects Agency, or DARPA), space (e.g. the National Aeronautics and Space Administration), and health (e.g. NIH) that have led to important inventions such as jet engines, radar, nuclear power, digital computers, the Global Positioning System (GPS), the Human Genome Project, and perhaps most significantly, the Internet (Janeway, 2012; Mazzucato, 2013; Gruber & Johnson, 2019). Successful examples of these require decentralization, active project selection (and a tolerance for failure), and organizational flexibility (e.g., Azoulay et al. 2018).

Climate change is a leading example of an area in which more innovation is needed to avoid environmental catastrophe, but where decentralized markets are unlikely to provide sufficient technology within the necessary timeline. It is important to remember that when the rate and direction of technological change is endogenous, horizontal policies like a carbon tax can be doubly effective because they reduce consumption of fossil fuels directly while also indirectly stimulating the development of clean technology. (Acemoglu, Aghion, Bursztyn, & Hemous, 2012; Aghion, Dechezleprêtre, Hemous, Martin, & Van Reenen, 2016). Despite this, it is clear that there are strong political obstacles to a carbon tax (or its equivalent, like “cap and trade”) that would be large enough to effectively combat global warming. The United States clearly needs to develop a portfolio of technologies to address climate change, and it needs a strategy to effectively deliver it.

### 3.2 Product Market Competition and Trade Policies

Industrial policy has earned a bad reputation because it has often involved heavy restrictions on competition, such as tariffs to protect infant industries from foreign
competition and relaxed antitrust policy to allow for more mergers to create national champions. The impact of competition on innovation is theoretically ambiguous. On the negative side, Schumpeter (1942) argued that the ex-post reward of innovation is monopoly profits, so increasing competition reduces incentives to innovate. On the positive side, monopolists have little incentive to innovate and replace the stream of rents they already enjoy, while new entrants are not similarly burdened (known as the “replacement effect” in Arrow, 1962). Existing empirical evidence suggests that competition typically increases innovation; especially if competition is initially low (see Van Reenen, 2011b for a survey).

There has been a great deal of research on the impact of trade with China on innovation over the last 20 years. China's growth as an export market is a clear benefit for innovation as it increases market size, which helps spread the fixed cost of R&D over a larger market (e.g., Grossman & Helpman, 1991; Bloom, Romer, Terry and Van Reenen, 2019). Much of this literature focuses on import shocks that increase competition, such as China's integration in the global market following its accession to the World Trade Organization in 2001. Shu and Steinwender (2018) summarize over 40 papers on trade and competition, arguing that in South America, Asia, and Europe, import competition mostly increases innovation (e.g., Blundell, Griffith, & Van Reenen, 1999; Bloom et al., 2016; Atkin, Khandelwal, & Osman, 2017). In North America, the impact of import competition is more mixed; for example, Autor, Dorn, Hanson, Pisano, and Shu (2017) find negative effects, whereas Gong and Xu (2017) find a zero effect.

In my view, the balance of the evidence suggests that greater trade competition typically increases innovation, and thus, current trade wars will be a detriment to growth. This conclusion means that industrial policies should be designed to encourage rather than chill trade competition (e.g., avoid protecting industries through high import tariffs). A better way is to encourage many entrants in areas of policy emphasis (e.g., environment) and award support that is based on merit. Moreover, policymakers must be prepared to allow many failures, which are inherent to experimentation, rather than assuming ex-ante that the government is capable of selecting winning approaches. The most successful industrial policies are based on this principle and include South Korean motor vehicles (Cherif & Hasanov, 2019) and the Taiwanese semiconductor industry that arose from Hsinchu Science Park (Chen, 2008).

4. Conclusion

Economic theory—and common sense—tells us that market economies will fail to provide a socially optimal amount of innovation. Reinvigorating technological leadership is not just a matter of national pride, it is necessary in order to sustain a robust middle class with good jobs at decent wages.
I have drawn on the most recent evidence to suggest three major areas where a largescale investment would have the greatest pay-offs: R&D tax credits, direct innovation grants, and expanding the supply of inventors (e.g., by relaxing skilled immigration rules). In my opinion, the largest, long-term effects would be through improving the opportunities of the many “Lost Marie Curies” and “Lost Einsteins,” talented, potential inventors who are held back by being born into disadvantaged backgrounds. Such a policy would reduce long-run inequality and increase growth, but would take many decades to have an effect. Therefore, a shorter-term program should also feature R&D taxes and subsidies.

Traditional approaches to industrial policy, which pick winners, are not desirable. However, the United States could learn from recent successes in East Asia and consider a mission-driven, industrial strategy in which the government creates a massive pool of R&D resources that are invested in the areas where market failures are the most substantial, such as climate change.

I propose the United States create a 10-year, $1 trillion Grand Innovation Challenge to reinvigorate R&D investment. At $100 billion per year (half of 1% of GDP), this program would still be less than half of the difference between federal R&D support today and that of 1964. If we are serious about building technological muscle back to the levels of the postwar period, we must make long-term investments that generate good, high-wage jobs.
References


PART III

INCREASING GOVERNMENT REDISTRIBUTION IN RESPONSE TO INCOME INEQUALITY AND DECLINING ECONOMIC MOBILITY

The Economics of Medicare for All
Craig Garthwaite

Universal Basic Income (UBI) as a Policy Response to Current Challenges
Melissa S. Kearney and Magne Mogstad

Wealth Taxation: An Overview of the Issues
Alan D. Viard

Policy Options for Taxing the Rich
Lily Batchelder and David Kamin
The Economics of Medicare for All

AUTHORS
Craig Garthwaite*

ABSTRACT
This memo provides a framework for evaluating the economic trade-offs of expanding the government’s role in financing and regulating health care through a single-payer system such as Medicare for All. First, I draw upon international comparisons to highlight important differences in the features of single-payer systems among high-income countries. I then discuss the economic trade-offs that would accompany the adoption of a single-payer system in the United States, including potential changes to the quality and quantity of medical services, and to the quantity of health-care products (such as pharmaceuticals). If such a system were to adopt the existing Medicare price schedule, the average quality of medical services would be expected to be lower, while the impact on the quantity of medical services would depend on the willingness of a single-payer monopsonist to exert downward pressure on the wages of health-care workers. A U.S. single payer could also exert its buying power to lower drug prices but doing so would likely reduce the future quantity of health-care products, since reducing the global profits of drug innovators would deter them from making the large investments in research and development (R&D) that are necessary to develop new products. Finally, I discuss market-based policy reforms that could promote affordability and access in the current U.S. health-care system.

* Northwestern University Kellogg School of Business, garthwaite@kellogg.northwestern.edu
1. Introduction

Current health-care policy debates reveal an unprecedented willingness among policymakers and voters to consider significantly expanding the government’s role in the health-care sector in the United States. The political popularity of Medicare for All is one such example. Historically, the term "Medicare for All" has been used to describe proposals that would expand Medicare coverage beyond qualifying elderly and disabled individuals to the entire nation, although there are now a wide variety of proposals that attempt to clarify how such an expansion would take place.

The common denominator of various Medicare for All proposals is an attempt to address two distinct but interrelated problems: The need to increase access to health insurance and the need to reduce health-care costs. Supporters of these proposals believe that a government-sponsored, single-payer system would address both of these challenges by providing universal access while also leveraging the government’s buying power to lower prices. However, these changes would undoubtedly have effects on a wide variety of outcomes, and policymakers should be aware of the potential unintended consequences of such an economically meaningful policy change. Rather than attempting to summarize the specifics of all the various proposals (all of which will undoubtedly change in the coming years), the purpose of this memo is to provide a framework for evaluating the economic trade-offs of expanding the government’s role in financing and regulating health care through a single-payer system such as Medicare for All.

Even after the passage of the Affordable Care Act (ACA), approximately 10% of Americans remain uninsured. While these individuals may still have access to emergency services funded by hospital uncompensated care (Garthwaite, Gross, & Notowidigdo, 2018), they lack the financial protection of health insurance and have difficulty accessing nonemergency services such as physicians and pharmaceuticals. In addition, rising deductibles and other forms of cost sharing have left millions of low-income individuals underinsured against the financial risk of negative health shocks.

The share of the United States that remains uninsured and underinsured is driven in part by the high cost of health care in the United States. While many policymakers blame high premiums on insurers’ profit margins, these premiums primarily reflect the prices of various providers and other firms in the health-care system. For this reason, a second goal of Medicare for All is to lower the cost of health care rather than merely subsidizing the cost of purchasing health insurance. The anticipated savings would primarily come from reducing administrative costs and expanding price regulation. In this memo I focus on the economic effects of expanding price regulation, since it is likely to be a larger source of potential savings and pose a

1 In thinking about this expansion, it is important to keep in mind that the government already has a meaningful presence in U.S. health care, accounting for over half of all spending.
greater potential disruption to the existing market of private providers and firms that serve at the center of U.S. health care. Because price regulation will not exist in a vacuum, we must consider the private sector’s response to such changes. These changes will alter the quality and composition of healthcare with meaningful economic costs and consequences. Thus, evaluating these reform proposals based on budgetary effects alone is at best incomplete and at worst disingenuous.

2. International Comparisons

Supporters of health-care reform in the United States often state that “every other developed country” has been able to achieve access to universal health care. This is a true statement, but it obscures the heterogeneity that exists across the health-care systems of developed countries.

In reality, developed countries use a variety of single- and multi-payer systems to achieve universal access to health insurance. Even single-payer systems can evolve in many different ways. Key differences across systems include eligibility criteria, models of cost sharing with patients, and the role of private health insurance (see Figure 1 for a full summary of the various decisions involved in designing a single-payer system). Each of these decisions will have a meaningful impact on the operation of a single-payer system in the United States.

Figure 2 contains a summary of how various developed countries have implemented these decisions. While there are many differences across these settings, one common feature across all systems is the government’s involvement in setting health-care prices.

Taken together, these figures make clear the diversity of universal health-care systems that actually exist across the developed world. Three features are particularly relevant when making comparisons to U.S. context: the role of private insurance firms, the role of private providers, and the price-setting mechanism.

Britain’s National Health Service (NHS) is often invoked during discussions about single-payer health care, but it is of little relevance as a point of comparison to the U.S. context. Not only does the British system provide universal access, it also features government ownership of facilities and employment of physicians. These features far exceed existing proposals in the United States, and the economic features of such a system are not easily compared to the U.S. context.

---

2 While administrative cost savings could exist, these will be small compared to the goals of the program. In addition, traditional fee-for-service (FFS) Medicare does not currently save money through a more efficient use of medical services. In fact, if we look at comparisons between Medicare Advantage (the private managed care version of Medicare) and FFS Medicare, we see that the private market is actually far better at providing incentives for lower utilization of health-care services (Curto, Einav, Finkelstein, Levin, & Bhattacharya, 2019; Baker, Bundorf, Devlin, & Kessler, 2016).
The Swiss system provides the closest comparison to a simple expansion of the existing Medicare program, in which 30% of enrollees receive their benefits through the privately administered Medicare Advantage program.\(^3\) The Swiss finance health-insurance purchases through a combination of government subsidies and private premiums. In turn, private firms compete for each citizen’s business. While many compare this system to the ACA exchanges, such a comparison misses an important difference: The Swiss system is based on a set of regulated prices for medical services. In that way, the Swiss system is more comparable to the Medicare Advantage market (the private managed care version of Medicare), where both explicit and implicit regulations allow Medicare Advantage prices to mirror those of FFS Medicare.\(^4\) This differs from the commercial market (both non-group plans such as the ACA and

---

3 Medicare Advantage (or Medicare Part C) is the private managed care form of Medicare that seniors can elect. Under MA, firms are paid a risk-adjusted payment for each enrollee and they are then responsible for all of their Medicare spending.

4 Explicitly, any provider that chooses to not enter an Medicare Advantage network can only charge a patient the FFS Medicare rate. This is vastly different from the commercial market, where out-of-network providers can effectively charge any price that they want. Implicitly, providers face a strategic dynamic where they know that if they attempt to charge too high of a price and Medicare Advantage providers can’t stay in the market, the enrollees will all simply default to FFS Medicare. So, the regulated price schedule stands as the outside option in the negotiations.
### Table 1: Key Features of Single-Payer Health-Care Systems in Selected Countries

<table>
<thead>
<tr>
<th>Design Features</th>
<th>Australia</th>
<th>Canada</th>
<th>Denmark</th>
<th>England</th>
<th>Sweden</th>
<th>Taiwan</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level of Administration</strong></td>
<td>National government</td>
<td>Provincial or territorial government</td>
<td>National government; administrative regions provide care</td>
<td>National government</td>
<td>National government; county councils responsible for most financing and purchasing</td>
<td>National government</td>
</tr>
<tr>
<td><strong>Eligibility</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Universal coverage</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Separate public programs for certain groups other than military</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Mandated Benefit Package</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hospital and physicians’ services</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Outpatient prescription drugs</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>LTSS</td>
<td>Limited</td>
<td>No</td>
<td>Yes</td>
<td>Yes, limited</td>
<td>Yes</td>
<td>Yes, limited</td>
</tr>
<tr>
<td>Dental, vision, and mental health services</td>
<td>Limited</td>
<td>No</td>
<td>Yes</td>
<td>Yes, limited</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Cost Sharing</strong></td>
<td>Yes</td>
<td>No</td>
<td>Yes, except visits without referrals</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hospital and physicians’ services</td>
<td>Yes</td>
<td>No</td>
<td>No, except visits without referrals</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Prescription drugs</td>
<td>Yes</td>
<td>n.a.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>LTSS</td>
<td>Yes</td>
<td>n.a.</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes, n.a.</td>
</tr>
<tr>
<td>Dental, vision, and mental health services</td>
<td>Yes</td>
<td>n.a.</td>
<td>Yes, for dental and vision</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Limit on out-of-pocket spending</td>
<td>No</td>
<td>No</td>
<td>No, but copayments decrease with higher out-of-pocket spending on prescription drugs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Reduction or exemption available</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Private Health Insurance</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Supplemental&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Substitute&lt;sup&gt;c&lt;/sup&gt;</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Other types of private insurance&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Participating Provider Rules</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Balance billing allowed</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Payments from private-pay patients for covered services</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Hospitals&lt;sup&gt;e&lt;/sup&gt;</strong></td>
<td>Mixed</td>
<td>Mixed</td>
<td>Public</td>
<td>Public DRG</td>
<td>Public Global budgets and DRG</td>
<td>Private FFS with overall global budget</td>
</tr>
<tr>
<td>Primary ownership</td>
<td>Mixed</td>
<td>Mixed</td>
<td>Public</td>
<td>Public DRG</td>
<td>Public Global budgets and DRG</td>
<td>Private FFS with overall global budget</td>
</tr>
<tr>
<td>Primary payment method</td>
<td>Global budgets and DRG in public hospitals; FFS in private hospitals</td>
<td>Global budget</td>
<td>Global budget</td>
<td>Global budget</td>
<td>Global budget and DRG</td>
<td>Private FFS with overall global budget</td>
</tr>
<tr>
<td><strong>Primary Care Physicians&lt;sup&gt;e&lt;/sup&gt;</strong></td>
<td>Private</td>
<td>Private</td>
<td>Private</td>
<td>Private</td>
<td>Mixed</td>
<td>Private</td>
</tr>
<tr>
<td>Primary employment</td>
<td>Private</td>
<td>Private</td>
<td>Private</td>
<td>Private Capitation</td>
<td>Mixed Capitation</td>
<td>Private FFS with overall global budget</td>
</tr>
<tr>
<td>Primary payment method</td>
<td>FFS</td>
<td>FFS</td>
<td>FFS</td>
<td>Capitation</td>
<td>Capitation</td>
<td>Capitation</td>
</tr>
<tr>
<td><strong>Outpatient Specialist Physicians&lt;sup&gt;e&lt;/sup&gt;</strong></td>
<td>Mixed</td>
<td>Private</td>
<td>Mixed</td>
<td>Public</td>
<td>Mixed</td>
<td>Private</td>
</tr>
<tr>
<td>Primary employment</td>
<td>Mixed</td>
<td>Private</td>
<td>Mixed FFS for self-employed providers; salary for public hospital employees</td>
<td>Public Salary</td>
<td>Mixed</td>
<td>Private</td>
</tr>
<tr>
<td>Primary payment method</td>
<td>FFS</td>
<td>FFS</td>
<td>FFS</td>
<td>Per-case payment</td>
<td>Salary</td>
<td>Salary</td>
</tr>
</tbody>
</table>

<sup>a</sup> Cost-sharing reductions or exemptions are available for prescription drugs in some provinces.

<sup>b</sup> Supplemental insurance could cover services not included in the single-payer plan, such as dental, vision, or hearing. It could also reduce enrollees' cost sharing, like the private plans that many Medicare beneficiaries purchase.

<sup>c</sup> Substitutive insurance, which duplicates the benefits of the single-payer health plan, could be offered to people who are not eligible for the single-payer system, such as noncitizens who have recently entered the country or temporary visitors. It could also be an alternative source of coverage if people are allowed to opt out of the single-payer system.

<sup>d</sup> Other types of private insurance could provide benefit enhancements, such as faster access to care, private rooms instead of semiprivate rooms for inpatient stays, and a greater choice of providers.

<sup>e</sup> Refers to the characteristics of a typical entity in each system.
employer-sponsored plans), where insurers often pay prices that far exceed those of FFS Medicare.

The Swiss system is not a true “single-payer” system as it involves multiple private firms that pay medical providers, but it does achieve universal access through a mixture of taxes and individual contributions. In contrast, the Canadian system provides the best comparison for a Medicare for All single-payer system, such as the one proposed by Senator and presidential candidate Bernie Sanders (I-Vermont). In the Canadian system, a single entity provides all health insurance, and residents are not allowed to purchase additional coverage for services that are already covered by the government insurer. Health-care spending is significantly lower in the Canadian system relative to the United States, which is a function of both lower prices for products and lower wages for providers. I will discuss both of these channels below and how they might inform the optimal structure for a U.S. single-payer system.

A distinguishing feature of the Canadian system is that it does not allow firms to offer coverage that enables individuals to “skip the line” or otherwise avoid the explicit rationing that is often inherent in single-payer systems. Patients who wish to do so must pay for such services entirely out of pocket from private providers. This feature

<table>
<thead>
<tr>
<th>Design Features</th>
<th>Australia</th>
<th>Canada</th>
<th>Denmark</th>
<th>England</th>
<th>Sweden</th>
<th>Taiwan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prescription Drugs</td>
<td>Internal reference pricing</td>
<td>External reference pricing</td>
<td>Internal reference pricing: price-cap agreement for drugs with no generic equivalents</td>
<td>Negotiated profit caps</td>
<td>Value-based payment</td>
<td>Value-based payment</td>
</tr>
<tr>
<td>Main Source of Financing</td>
<td>General tax revenues and earmarked tax revenues</td>
<td>Provincial and federal general tax revenues</td>
<td>Earmarked income tax</td>
<td>General revenues and payroll taxes</td>
<td>General revenues raised by county councils, municipalities, and nationally</td>
<td>Payroll-based premium, supplementary premium based on nonpayroll income, general revenues, tobacco tax, lottery gains</td>
</tr>
</tbody>
</table>

DRG = diagnosis-related groups; FFS = fee for service; LTSS = long-term services and supports; n.a. = not applicable.

a. Cost-sharing reductions or exemptions are available for prescription drugs in some provinces.
b. Supplemental insurance could cover services not included in the single-payer plan, such as dental, vision, or hearing. It could also reduce enrollees’ cost sharing, like the private plans that many Medicare beneficiaries purchase.
c. Substitutive insurance, which duplicates the benefits of the single-payer health plan, could be offered to people who are not eligible for the single-payer system, such as noncitizens who have recently entered the country or temporary visitors. It could also be an alternative source of coverage if people are allowed to opt out of the single-payer system.
d. Other types of private insurance could provide benefit enhancements, such as faster access to care, private rooms instead of semiprivate rooms for inpatient stays, and a greater choice of providers.
e. Refers to the characteristics of a typical entity in each system.

Source: Congressional Budget Office, 2019
greatly limits the scope of services that can exist outside of the government insurance system. Allowing individuals to purchase additional private insurance can help to mitigate a significant downside of regulated prices—the reduced availability of high-quality options even for those who are willing to pay—though this will ultimately depend on how many people opt out of the public insurance system by purchasing private coverage and the fixed costs of operating private facilities.  

3. The Economic Trade-Offs of Single-Payer Health Care in the United States

A single-payer system leverages the buying power of the single buyer to hold prices below the market outcome. However, such a massive change to U.S. health-care policy would affect many different levers in the health-care system. Those levers could, in turn, influence the new equilibrium price of health care. Many analyses that are favorable toward a single-payer system are incomplete because they rely upon a “partial equilibrium” analysis. That is, they ask the question: If we hold everything else constant in the system, what happens if we increase the use of Medicare’s regulated price schedule?

In reality, all else will not remain constant if such a significant policy change is made. Each actor in the system will re-optimize and create a new equilibrium of prices, quality, and quantities, which would affect current Medicare enrollees, current health-care providers, potential future providers, and new potential enrollees to the program. Therefore, the economist’s job is to make an informed prediction about the final outcome of a policy change on prices once all the components of the system have adjusted to the new equilibrium.

Savings from a Medicare for All system would come from many categories. Some of the savings would come from a change in the nature of administrative costs in the system. Changing the nature of administrative costs will have a number of economic effects, with the sign and magnitude of some effects currently ambiguous. First, in a true single-payer system, providers would need to expend fewer resources to comply with insurers’ systems, likely reducing costs.

Second, a government single payer would not need to advertise for potential customers, which could greatly reduce expenditures in this category. However, to the extent that advertising and competition provides incentives for differentiation and innovation—particularly with respect to innovations that attempt to limit the moral hazard inherent to health insurance—the magnitude of the economic (as opposed to the accounting) savings are less clear.

5 While a facility would not set its price based on the fixed cost, the very emergence of certain providers would depend on whether they believed in the long run their variable profits would exceed the fixed costs of entry.
The final contributor to Medicare's lower administrative costs is the system's relatively lax approach to utilization management. Broadly speaking, Medicare does little to control the quantity of medical services that enrollees use. The economic savings of moving all enrollees to such a freewheeling system is unclear. While the systems of prior authorization, step therapy, and other attempts to regulate moral hazard utilized by insurers carry meaningful costs, they have the potential to decrease the use of inappropriate and cost-ineffective care. The amount of utilization management under a proposed single-payer system is unclear, but would greatly affect people's interaction with the new system.6

The vast majority of proposed savings from a single-payer system would come from the expanded use of regulated prices for providers and other medical services. The common characteristic of universal health-care systems across the developed world is a willingness to exploit the government's buying power, which brings prices below the market outcome and also impacts the optimal quantity and quality of medical services.

A single-payer system in the United States is likely to substantially reduce payments for medical services, including those to physicians and facilities such as hospital and outpatient clinics. The potential scope for price reductions is quite large because most commercial insurers pay rates that are well in excess of those charged to Medicare. For example, a recent RAND Corporation study found that for a sample of hospitals in 25 states, the average hospital charged private insurers 240% more than Medicare rates (White & Whaley, 2019).

Because a single-payer system would grant the government monopsony power in the labor market for health-care workers, policymakers should consider how workers and suppliers would respond to greater price regulation. The response is likely to vary based on the specific markets for different products and services. Firms could adjust quantity, quality, or both. Hospitals, for example, could decrease the use of private rooms, substitute labor for lower-cost sources (i.e. more mid-level providers and fewer MDs), etc. In addition, to the extent the government applies its monopsony power to products such as pharmaceuticals, it will also impact the incentives for private firms to invest capital in the development of new products.

In the next section, I present empirical evidence about the effect of regulated prices on the quality of medical services, the quantity of medical services, and the quantity

6 Describing this point, the Congressional Budget Office (CBO) notes: “In the United States, public programs have implemented few utilization management programs, but private insurers have increasingly used them to lower costs. Some private insurers require prior authorization for patients seeking expensive therapies, for example, and Medicare Part D plans offer low or no copayments to patients who use cheaper generic medications. Many of those strategies could be continued under a single-payer system. The utilization management in such a system might not be much of a change for people who were previously enrolled in a private plan, but it would impose new constraints on the choice of health care services for those who were previously enrolled in the Medicare FFS program.”
of products. While this research cannot provide definitive evidence about the exact magnitude of the effect of government buyer power, they provide evidence about the nature of such effects.

3.1 Changes in the Quality of Medical Services Under a Single-Payer System

How will the quality of medical services provided by a strategic firm change in response to the introduction of a single-payer system? Most of today’s Medicare recipients are quite happy with the quality of services they receive, and despite the use of regulated prices, they enjoy access to a wide range of high-quality hospitals. But this is not predictive of a beneficiary’s satisfaction under a new program, nor of existing enrollees’ experience if the program is expanded.

A new equilibrium in which all hospitals earn only the current Medicare reimbursement would result in a very different experience for Medicare recipients. Hospitals serve a broad swath of patients. These patients each pay different prices for the services they receive, but they typically fall into three groups: Medicaid recipients (18.5% of 2016 revenue), Medicare recipients (40.8% of 2016 revenue), and the privately insured (33.4% of 2016 revenue). Medicaid pays a regulated price that is thought to approximate marginal costs; Medicare pays a regulated price that is thought to approximate the average costs of the average hospital; and privately insured patients pay a negotiated price that reflects the relative bargaining power of the provider and the insurer and which is usually much higher than those paid by Medicaid and Medicare recipients.

Now, consider the decision of a hospital about how and whether to invest in quality. A hospital could make costly investments in quality in an attempt to attract patients. Doing so, however, would greatly reduce its Medicare margin, which is driven by the cost structure of the average hospital across the country. Effectively, the choice by a hospital to invest in costlier quality than the average hospital lowers the margin they can earn from a public payer because that payer does not respond to an individual hospital’s strategic investment but rather to the decisions of the average hospital.

This is not true in the private insurance market, where if costly investments in quality increase patients’ willingness to pay, they could also increase the negotiated rate between hospitals and insurers. Thus, hospitals will make investments in quality to

---

7 Some proposals for Medicare for All recognize this point and therefore propose setting reimbursement at some multiple of Medicare. However, it is unclear (a) whether this multiple is sufficient, and (b) whether this multiple will sufficiently change over time.

8 Note that while people generally refer to this as “cost shifting”—hospitals charge the privately insured more because the public sector doesn’t pay enough—this is simply a form of price discrimination where firms charge based on a patient’s willingness to pay.

9 This is primarily true for the FFS forms of public insurance. Both Medicare Advantage and Medicaid Managed Care tailor rates to individual hospitals, but not nearly to the same degree as the commercial market.
attract privately insured patients if they believe that the return from that investment will exceed the lost profits from the lower margin that they earn on Medicare patients.

Understanding this trade-off between quality and margin is critical for understanding the trade-offs for any Medicare for All-style plan. While some have suggested that regulated prices would merely trim away profits and unnecessary services, it is far more likely that hospitals will have to make meaningful changes to the quality of services that patients in the private market are currently willing to pay for (via higher premiums). The change in quality is fundamentally related to the reimbursement rate set by the single payer—if a future single payer were to pay a higher rate, the quality declines could be mitigated. However, there would be limited ability of hospital quality to vary to the same degree as consumer preferences. In addition, these higher rates would decrease the potential savings from such a system.

The decline in overall quality in exchange for expanded coverage and reduced prices might be an optimal decision from the point of view of society. This, however, is ultimately the debate that we should be having, rather than suggesting that the only losses from a single-payer system will be profits and inefficiency.

### 3.2 Changes in the Quantity of Medical Services Under a Single-Payer System

Approximately 60% of health-care spending goes to labor costs. Any attempt to reduce spending through lower prices will ultimately affect the wages of medical providers. There are simply not enough profits in the system to generate the type of savings that would be required for all providers to operate under existing Medicare reimbursements and still earn the same wage.

The economic costs of using market power to reduce these wages depends on the responsiveness of medical providers to lower wages. Some providers may decide to substitute leisure for work, or switch to a different occupation. Fewer individuals may undertake the training necessary to become medical providers in the future (if you reduce the returns to medical school, fewer people will attempt to become physicians and will instead go into other sectors). A large reduction in wages is unlikely to meaningfully lower the absolute quantity of physicians, since medical schools artificially constrain the number of student slots they offer. However, a decrease in the number of applicants could reduce the average quality of physicians entering the market since medical schools make admissions decisions based on their assessment of the quality of the marginal applicant.

While the effects of a single-payer monopsonist are hard to predict, Canada’s experience and decisions regarding physicians’ salaries may be informative.10 In

---

10 This would be a system where the government serves as the only insurer and individuals are not allowed to supplement payments to providers with additional insurance.
Canada, provincial governments offer insurance; care is provided at privately owned facilities by privately employed physicians, which is similar to how I would expect a single-payer system to operate in the United States. Canadian physicians earn lower salaries compared to those in the United States, however it is unclear the extent to which this reflects the single payer’s monopsony power or broader labor-market differences across the two countries. I examine this question in ongoing work with coauthors, by comparing the distribution of wages for health and non health workers across the two countries (Chown, Dranove, Garthwaite, & Keener, 2019).

If the lower health-care wages in Canada are the result of buyer power, then we would expect workers with options outside of the health-care market to earn similar wages across the two countries. This would include, for example, unskilled health-care workers who could easily leave health care for another sector. If lower physician wages in Canada resulted from monopsony power, we would expect the wage differences for highly skilled health-care workers in Canada and the United States to be greater compared to the wage differences among workers in other sectors of the economy, with the magnitude of this wage difference reflecting the expression of buyer power by the government insurer.

Figures 2 and 3 provide some evidence that the Canadian monopsonist does not meaningfully exert its market power on physician wages. Instead, a large proportion of the difference in provider wages across the two countries reflects other labor-market differences, such as the general wages earned by highly trained professionals in the market. Wages for lower skill employees across health care and other sectors are quite similar, suggesting the monopsonist is not using its buyer power to push down wages of lower skill health-care workers.

Among high-income workers, those in the United States earn more than those in Canada. However, this difference also holds in other sectors throughout the economy and reflects the high wages earned by those at the top of the U.S. income distribution. Chown et al. (2019) estimate highly educated Canadian health-care workers earn 26% less than those in the United States. If we look at similarly educated workers outside of health care, they earn 22% less in Canada than they do in the United States.

These wage differences suggest the Canadian monopsonist does not meaningfully exert its massive buyer power on the wages of health care workers. One reason why a monopsonist would choose not to exert such power is if it were worried that supply of a good (in this case physician labor) is fairly elastic and thus a decline in wages would meaningful deceases the quality or quantity of health-care providers.11 This suggests that a similarly situated U.S. monopsonist may have limited scope to reduce spending by suppressing provider wages.

---

11 It’s possible that a health-care monopsonist could exploit altruistic motives of physicians to charge below the market wage, but there is likely a limit to the size of this effect.
Figure 2. Income Distribution by Country

Notes: Figure plots the mean income for each within-country income ventile for all workers and workers in a health-care occupation. For Panel B, income quantiles are calculated only among individuals in health-care occupations. Sample is employed, paid workers in the 2011 National Household Survey (Canada) and 2010 American Community Survey (U.S.), using the harmonized versions of each data source from IPUMS International.
Figure 3. Income Distribution by Country for Advanced Degree Holders

Notes: Figure plots the mean income for each income ventile (within country and degree status) for all workers and workers in a health-care occupation. For Panel B, income quantiles are calculated only among individuals in health-care occupations. Sample is employed, paid workers in the 2011 National Household Survey (Canada) and 2010 American Community Survey (U.S.), using the harmonized versions of each data source from IPUMS International.
While a U.S. single payer could choose to exert more monopsony power than its Canadian counterpart, doing so would involve a different set of economic considerations than is often assumed by those who are concerned about the difference in provider wages across the two countries. These are costs that it appears the Canadian monopsonist is unwilling to incur, based on their revealed preferences.

### 3.3 Changes in Product Quantity Under a Single-Payer System

A single payer's ability to negotiate pharmaceutical prices is another source of significant potential savings. It is often claimed that Medicare does not currently negotiate pharmaceutical prices. While it is true the Center for Medicare and Medicaid Services (CMS) does not directly negotiate prices, private firms operating under Medicare Part D and Medicare Advantage do negotiate the prices of retail pharmaceutical products. These organizations are quite skilled at negotiations and their bargaining power is strong, so it is not clear CMS would earn a larger discount if it were to negotiate directly.

However, for physician-administered drugs (i.e., those covered by the Medicare Part B program), Medicare does not negotiate any price concessions. Instead, the government has formally established itself as a price taker where they pay a fixed markup over the average price in the private market. There is certainly more room for negotiation for these products.

A single-payer would again leverage its buyer power in the pharmaceutical market. Table 2 shows the average price paid for a comparable basket of drugs across the United States and Canada. The Canadian monopsonist does appear to exploit its position in the market for pharmaceuticals far more than it does in the health-care labor market. For example, Chown et al. (2019) finds that for a comparable set of drugs, Canadian consumers pay approximately 54% less than patients in the United States. This difference is even greater for non-neurological drugs, which likely reflects the heavy consumption of neurological drugs by Medicaid patients in our sample. While some of the difference in prices could result from other differences between the two markets, there are very few differences between U.S. and Canadian prices of non-health-care products.

---

12 Price data for the United States is backed out of a sample of Medicaid drugs—and therefore accounts to some degree for the existence of rebates. Price data for Canada comes from the province of Ontario.

13 This, in fact, could be an underestimate of the difference given difficulty fully estimating the magnitude of rebates. We use data from Medicaid rebates, but for many reasons this could distort estimated difference.
Table 2. Prescription Drug Price Indices

<table>
<thead>
<tr>
<th></th>
<th>Estimated Price Index (Canada relative to United States)</th>
<th>Share of Medicaid Spending in Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sample</td>
<td>0.46</td>
<td>0.26</td>
</tr>
<tr>
<td>Neurological Drugs</td>
<td>0.48</td>
<td>0.36</td>
</tr>
<tr>
<td>Non-Neurological Drugs</td>
<td>0.36</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Notes: Table shows estimated price indices for the sample of brand name prescription drugs described in Chown et al. (2019). Share of Medicaid Spending in Sample gives the fraction of Medicaid spending on prescription tablets and capsules in the first quarter of 2018 reflected in our estimation sample for each drug class group.

The Canadian monopsonist appears to exercise its buyer power when it is optimal to do so. But how is Canada able to exert this power? And why is it optimal for Canada to exert its buyer power in the product market but not the labor market?

A large single payer can extract lower prices not only because of its size but also because of its willingness to walk away from a negotiation if it does not receive a satisfactory price. Although Medicare is a large buyer, it is required to supply nearly all drugs (this is particularly true in the case of Medicare Part B). For this reason, the CBO has estimated that allowing negotiation is unlikely to change prices (Congressional Budget Office, 2007). However, this analysis is based on a scenario in which Medicare has the authority to negotiate but not to deprive access to Medicare enrollees when monopoly prices are too high (known as a closed formulary). If Medicare had this additional leverage in pricing negotiations, it could almost certainly lower prices (particularly on Part B drugs). However, there could be a political cost if the government deprived seniors of access to some medications solely because of their price.

The comparison between Canada and the United States in the product market is less apt. In a global product market, a single payer must consider how exercising its buying power will impact the producer’s future global profits and not simply the price of that product in the domestic market. After all, a monopsonist wishes to avoid reducing the incentives to develop new products, and producers make this decision not based on any one country but instead on the expected global profits. Given the relative market share of Canada and the United States in the global market, the two countries are likely to reach different conclusions. Because the United States accounts for a larger share of the global market, its pricing decisions have far more influence on the pace of development of future products.

To understand the potential effect of buyer power in the health-care product market, consider the strategic decisions of pharmaceutical firms, which make large, risky investments in research and development (R&D). The patent system rewards
innovative firms by granting them a temporary monopoly, which allows innovators to recoup their investment before competitors enter the market.\footnote{While the product is under patent, no firm can make a product containing the exact chemical composition. However, a competitor can make a therapeutic substitute that targets the same condition and even uses the same mechanism of action as long as the product is sufficiently different in composition to secure a unique patent.}

The pharmaceutical industry is characterized by high fixed costs (R&D) and low marginal costs of production.\footnote{This is clearly true about small molecule products. As the industry as evolved to produce more biologic products and now with more gene therapy products the marginal costs of production have grown.} On the margin, firms will earn profits even at relatively low drug prices once up-front investments are made. Thus a U.S. monopsonist single payer could exert market power to lower prices without scaring away existing pharmaceutical producers. However, lower prices are more likely to deter firms from making large investments in R&D to develop new products. For this reason, a wide body of literature shows a robust connection between market size and investment in R&D (Finkelstein, 2004; Acemoglu & Linn, 2004; Blume-Kohout & Sood, 2013; DuBois, de Mouzon, Scott-Morton, & Seabright, 2015).

The pharmaceutical market illustrates the trade-off between two forms of inefficiency. Governments allow the \textit{static inefficiency} of monopoly prices vis-à-vis patenting in order to avoid the dynamic inefficiency of reduced innovation in the future. Pharmaceutical manufacturers make R&D investments based on the potential global profits. This provides an opportunity for a relatively small country such as Canada (which has fewer than 40 million residents) to choose to exercise buyer power without meaningfully reducing the development of future products. The Canadian monopsonist faces a much lower elasticity of supply of future products.

A larger country faces a higher elasticity of supply for new products because its citizens make up a larger share of the global market. Therefore, its decisions will have a greater impact on global profits. A larger single payer will be less likely to exert the same degree of market power compared to smaller counterparts, since its decision to exercise buyer power will require the sacrifice of future pharmaceutical innovation. Thus, citizens must decide how much they value drug innovation versus low drug prices. This is a very fair debate to have, but it is far more nuanced than a simple discussion about whether the United States should pay lower prices for drugs.

4. \textbf{How Should We Think About the Cost Estimates of a Single-Payer System?}

Just as there are numerous versions of Medicare for All, so too are there a plethora of cost estimates available for a potential single-payer system. At this preliminary point, sorting through these estimates does not serve a lot of value. Instead, it is important
to consider what the appropriate methodology would be for evaluating the potential costs associated with any increase in government-provided health care.

An increase in the size of the Medicare population will increase government spending, but does this spending represent net new outlays, or is it simply a shift from private premiums to public dollars? If a new enrollee previously had insurance through her employer, federal spending on her insurance will supplant premiums that she previously paid. As a result, the wage portion of her compensation will increase (Baicker & Chandra, 2006; Gruber, 1994). In turn, tax revenues will also increase because the compensation that was previously provided as health insurance was tax deductible. At a cost of approximately $280 billion per year, the tax deductibility of health insurance is the single largest expenditure in the tax code (Tax Policy Center, 2016).

If individuals are receiving government insurance which reduces their premium expenditures, then the government has the ability to raise taxes without harming economic performance. The distributional implications of such a tax are more complicated since the distribution of the current burden of health insurance premiums for individuals does not necessarily match the distribution of the current burden of their income taxes. This is especially true for wealthy individuals who spend a much greater share of their income on taxes than they do on health insurance premiums. Therefore, if the argument is that new taxes simply reflect existing payments for health insurance premiums (and therefore have minimal negative economic consequences), utilizing a broad-based payroll tax may be preferable to using the existing income tax structure to fund these new government expenditures.

An additional question for consideration is whether individuals will choose to purchase additional insurance coverage (assuming it is legal to do so). To the extent this is an issue, the increase in individual incomes resulting from the introduction of a Medicare for All system would be muted; thus, tax increases would pose greater economic costs.

5. Market-Based Policies to Improve U.S. Health Care

While the costs of Medicare for All could be substantial, the goals of expanding coverage and lowering costs are laudable and should be a policy goal supported by all. Instead of promoting a complete overhaul of the U.S. system, which will likely throw out the good with the bad, I argue the United States should strive for a more modest goal of restoring competition to parts of the health-care market where it is currently lacking.

Admittedly, the package of policies discussed below does not have the “home run” quality of a single, large policy that will “solve” the problem of U.S. health care, but such home runs will most likely cause more harm than good. The U.S. health-care
system accounts for approximately 18% of the U.S. Gross Domestic Product (GDP), placing its size roughly on par with the entire economy of Germany. A system of this magnitude cannot be disrupted overnight. Instead, I argue that policymakers should look for incremental approaches to promote competition in all sectors of health care. If this is accomplished, prices will decline to a competitive level, which will allow our existing social insurance programs (i.e., the ACA marketplaces, Medicaid, and Medicare) to expand coverage.

I will discuss a number of such policies—but will also note that this list is not meant to be comprehensive.\textsuperscript{16} Instead, these are examples of the types of focused policies that we should be pursuing. At a high level, we can break these policies into those that make the overall health-care market more efficient and those that strive to make Medicare more efficient.

### 5.1 Improve Overall Competition in Health-Care Markets

Unlike other developed countries, the United States relies heavily on the private market to finance and provide health-care services for its citizens. There are many advantages to a market-based health-care system. The citizens of a large and diverse country such as the United States will have a wide variety of preferences and meaningful differences in their willingness to pay for health-care quality. Regulated prices and central planning (by either a government entity or an independent third party) are unlikely to maximize welfare, and the market can more efficiently allocate goods and services. This is especially true considering the large number of economic actors involved in developing innovative new health-care products and services. It is hard to imagine what omniscient actor could more efficiently balance these forces. For this reason, despite many contentions to the contrary, an appropriately regulated, market-based system remains the best mechanism for maximizing welfare.

However, relying on the market for the provision of such a vital set of goods and services requires policymakers to recognize that health-care markets, like any other market, can fail. Market structures and institutions must be vigilantly protected in order to promote robust and vigorous competition. Unfortunately, there are many areas of the U.S. system where competition is not being fostered and government policy is actually undermining market competition. I address some of these concerns below.

#### 5.1.1 Promote Generic Competition

U.S. pharmaceutical policy has sought to balance innovation and affordability by granting innovating firms with a new product a temporary period of market exclusivity before generic competitors may enter the market. However, over time, brand-name

\textsuperscript{16} A more complete discussion of these points can be found at: \url{https://www.kellogg.northwestern.edu/faculty/garthwaite/htm/Garthwaite_Testimony_Judiciary_Final.pdf}
drug manufacturers have found ways to deter generic manufacturers from bringing competing lower priced products to market. In addition, some fundamental market structures, such as small market generics, limit the existence of multiple competitors and allow firms without patent protection to effectively act as monopolists and earn excessively high price-cost margins. I will discuss each of these factors in turn.

First, we must lower the barriers to entry for generic drug makers once the patent protection of an innovative firm has ended. Policymakers should ensure that potential generic entrants have an opportunity to demonstrate their product’s bioequivalence to a patented product. Unfortunately, some brand-name manufacturers go to great lengths to restrict access to their product so that generic firms cannot accumulate enough samples of the brand-name drug to demonstrate a generic drug’s bioequivalence. Brand-name firms often do this by abusing regulations that are intended to promote the safety and security of the pharmaceutical supply chain. This should be illegal. The pending Creating and Restoring Equal Access to Equivalent Samples Act (CREATES) would accomplish this by requiring firms to make such samples available.

While lowering entry barriers should be a primary goal, we also must confront the fact that there are a number of generic markets where the target population is so small the market will never support multiple competitors. In a well-functioning generic market, firms compete primarily on price. Profits therefore are determined by a firm’s ability to manufacture products at the lowest marginal cost. This fierce price competition means that successful entrants must be able to produce enough to reach the minimum efficient scale (MES) of their production process (i.e. the quantity at which the costs of production are minimized). Absent sufficient quantity, entrants realize they will find themselves at a perpetual cost disadvantage to incumbent firms and therefore will rationally decline to enter the market. For sufficiently small markets, there is only enough demand for a single manufacturer to reach MES—and the incumbent firm is a natural monopolist that maintains meaningful pricing power.

In recent years, cognizant of the pricing power available to manufacturers of generic products with sufficiently small potential markets, a number of firms have adopted a strategy of acquiring small-market generics and significantly raising prices (Hopkins & Martin, 2018; Pollack, 2015; Rockoff & Silverman, 2015). These cases are not examples of the above-discussed trade-off between access today and innovation tomorrow—society has long since paid for the innovation from any of these products. Instead, the high prices represent firms taking advantage of a market failure created by the small patient population.17

17 While large pharmaceutical firms were historically either unwilling to exploit this pricing power or unaware of this financial strategy, the practice of firms charging high prices without fear of entry in small generic markets is now widespread throughout the industry (albeit the strategy is typically employed by smaller firms with fewer invested assets in the industry).
I propose the Food and Drug Administration (FDA) be required to identify markets that appear to be natural monopolies and then undertake a request for proposal (RFP) process for those markets. Under this RFP process, any private firm could apply for the rights to be the exclusive manufacturer of a natural monopoly generic medicine at a certain fixed percentage above manufacturing costs; firms would compete on the amount of margin they would in order to require to serve the market. The winning firm would possess the exclusive rights to sell the drug at this regulated price for a time period sufficient to recover the fixed costs of entry. At that time, the FDA would have the option of re-auctioning off this new form of market exclusivity.18

Recent scientific advances have allowed for an increasing personalization of medicine. Along with coauthors, I have documented the rising share of clinical trials involving a patient-specific biomarkers to determine either efficacy or safety (Chandra, Garthwaite, & Stern, 2018). Almost by definition, personalized medicine will involve products with limited patient populations, and for many of these products we should be worried about whether robust generic competition will ever emerge.19 While the problem of small-market generics is not a dominant feature of today’s market, it will only grow in importance. It will likely be easier to address the problem now than it will be when the number of powerful interests manufacturing such products increases.

5.1.2 Improve Biosimilar Adopts by Regulated Contractual Form of Rebates

Price negotiation in pharmaceuticals occurs through the use of rebates, which are discounts off of the listed price that is negotiated between the pharmacy benefits managers (PBMs)20 and pharmaceutical manufacturers. Manufacturers are willing to give larger rebates when a PBM can credibly signal that they will shift a large volume of sales toward their product. One way that PBMs do this is by promising the manufacturer that their product will have the lowest cost sharing (i.e., copayment or coinsurance) among all its potential competitors in a therapeutic class. This is accomplished through various tiers of a formulary (the list of drugs and cost sharing for consumers).

Many contracts specifically reference potential rival products that might serve as a competitor—they specifically state that other competitors not be on a particular

---

18 In order to ensure the efficient operation of this process, it may also be necessary for the FDA to set a maximum percentage that they will accept before they will turn to a nonprofit or government supplier for the product. This will limit any ability of firms to collude to divide up the markets in which they choose to enter.

19 The problem of competition for precision medicine will be further complicated in situations where the patented product is a biologic.

20 Pharmacy Benefit Managers are private firms that manage the prescription drug portion of an individual’s health insurance benefit. This involves a number of tasks, but perhaps most important is the negotiation of drug prices with pharmaceutical manufacturers through a system of confidential discounts (i.e., rebates) from a publicly known list price.
formulary tier. These contracts that reference a rival can either be pro- or anticompetitive depending on the economic context. If there are a large number of products in the market and patients can be easily moved across products (such as in the small molecule market), then these contracts likely improve efficiency.

However, patients are unwilling to move across some types of products (and we may not want them to switch across products for medical reasons). In those settings, particularly if the incumbent firm has a large market share, rebate contracts that reference a rival can be anticompetitive. This is because a potential entrant can only compete for treatment-naive patients (i.e. those that have not previously successfully used one of the treatments). Therefore, if the rebate for the entire patient population is contingent on the entrant not being on the preferred tier, there is no price the entrant can offer that would be worth more than the rebate on the stock of patients that have already been using the drug. In such settings, we need to more carefully evaluate whether contracts that reference rivals are anticompetitive.

5.1.3 More Complete Review of Potentially Anticompetitive, Hospital-Insurer Contracting

There is a great deal of attention paid to the prices of pharmaceuticals relative to the share of health-care spending (15% to 20%) they comprise. Relatively less attention is paid to the prices charged by hospitals and other medical providers, which comprise a much greater share of health-care spending. Some of these high prices, particularly for hospitals, are the result of quality and brand preferences across consumers. However, we are increasingly worried that some of these prices are the result of hospital consolidation and selective contracting.

In particular, there are concerns that large health systems are exploiting their market power to require contracts that inflate prices across all hospitals in the system. These include contracts that reference rivals, most favored nation clauses, and anti-tiering/steering contracts that require all facilities in a system to be on the most preferential network tier in order for any to be on that tier. Given the increasing prevalence of large health systems, it is important that competition authorities undertake vigorous review of these contracts. In addition, it is important that this review extend to nonprofit hospitals and health systems. Currently, the Federal Trade Commission (FTC) is limited in its ability to regulate these providers (outside of merger review), which stands at odds with the evidence that many nonprofit hospitals appear to act similarly to their for-profit counterparts (Dranove, Garthwaite, & Ody, 2017).

---

21 Formularies are lists of drugs with tiers based on the cost sharing patients must pay to access the drug. For example, a formulary could have three tiers with a generic tier having a copayment of $5, a “preferred brand” tier with a copayment of $15, and a “nonpreferred brand” tier of $25. Cost sharing can also be based on a percentage of the drugs cost (i.e. coinsurance), which is often the case for expensive specialty drugs. Higher cost sharing decreases utilization, and thus manufacturers attempt to gain access to lower formulary tiers by offering larger price discounts.
5.2 Improve Efficiency of Medicare

There are also incremental changes that could be made to Medicare that would promote competition and improve efficiency. These changes are important for two reasons. First, demographic change in the United States will increase the importance of Medicare to both the federal budget and the health-care sector. Second, to the extent that Medicare becomes the vehicle for greater health-care coverage, improving its efficiency is a useful policy goal.

5.2.1 Improve Competition for Pharmaceuticals Purchased by Medicare Part D

Medicare Part D is an explicit public-private partnership in health care where the government subsidizes the purchase of insurance but the development and offering of plans is undertaken by private firms. When the program was created, its goal was to use market forces to promote competition and efficiency.

There are several features of Part D that subvert this goal. The first is Medicare Part D’s reinsurance program, which shields private firms from the cost of very expensive drugs. After an enrollee spends about $5,100 in out-of-pocket spending on drugs, they enter the “catastrophic coverage” range in which the government is responsible for 80% of costs, firms for 15%, and enrollees for 5%. Therefore, private firms have little incentive to engage in price negotiations for the most expensive drugs. Perhaps more concerning, PBMs operating in both the commercial and the Part D markets may face different incentives for rebates across these different markets and could use the confidential nature of rebates to increase government Part D spending.

Reinsurance may have been necessary to initially attract firms to the market at the program’s inception. Now that participation in Part D is well established and quite profitable for firms, the reinsurance program is no longer necessary. Therefore, I propose that Congress either end the reinsurance program entirely or greatly curtail its generosity so that plans are responsible for 80% of costs and the government is only responsible for 15%.

A second feature of Part D that decreases competition (and might affect prices) is the institution of “protected classes,” which require firms to cover all products in six protected therapeutic areas (immunosuppressants, antidepressants, antipsychotics, anticonvulsants, antiretrovirals, and antineoplastics). Limiting the formulary makes it very hard for plans to negotiate large discounts and may shift investments in drugs toward these classes. While it is clear that we need to balance the trade-off between price and access when we consider optimal formulary design, the current system errs too far on the side of access. Therefore, Congress should consider amending protected classes to allow more utilization management for these drugs.
5.2.2 Introduce More Competition to Medicare Part B

While Medicare Part D has an established structure for negotiating pharmaceutical prices, physician-administered drugs are covered under Medicare Part B and involve no negotiation at all. Instead, Medicare pays for these products on a cost-plus basis (physicians purchase the products and are then paid the average price plus a 4.3% margin). This perverse system creates an incentive for firms to charge higher prices in the private market and for physicians to prescribe drugs with higher prices.

Given the growing importance of physician-administered drugs, a category of products that include oncology products, it is essential that Medicare introduces some competitive pressure into the pricing of Part B drugs. While some have called for covering all products under Medicare Part D, doing so would likely expose many patients to more onerous cost sharing than they currently experience. Therefore, a better potential solution is to create the structures for PBM-like vendors to emerge and handle the negotiation for these products. This would require physicians to no longer take financial title to these products in the first place, which would eliminate the incentive to prescribe more expensive drugs without exposing them to carrying costs associated with the most expensive products.

Some providers may argue that the funds they currently receive for Part B provide reimbursement for other valuable medical services that they provide. This could be particularly true for some safety-net providers. However, to the extent that this is true, we should directly pay providers for these services rather than continue with a system that raises prices in part of the market in a Rube Goldberg-like attempt to finance other parts of the system.

5.2.3 Fixed Risk Adjustment in Medicare Advantage

The greater use of private providers in the Medicare Advantage program introduces a tension between providing strong incentives for cost controls and ensuring that individuals with high medical expenses receive appropriate access. On the one hand, the very purpose of privatizing this benefit is to provide a firm that is the residual claimant on health spending (i.e., they keep what is not spent) with the incentive to control costs. On the other hand, this creates strong incentives for firms to serve only healthy applicants who naturally have lower health-care costs with any effort from the firm.

To address this concern, Medicare Advantage program payments to providers are adjusted for the risk of the patient. For each patient, firms submit diagnostic codes that are used to calculate a patient-specific risk score. The expected spending for each risk score is derived from the spending by people with similar scores in the FFS system.

Under ideal settings, this would result in firms having the incentive to attract sicker patients and then actually manage their risk. Unfortunately, in reality, private plans
have a strong incentive to maximize the risk scores of enrollees by costly activities, such as reviewing the medical charts to provide support for additional diagnoses. At the extreme, this could lead to “upcoding,” or the inclusion of inaccurate risk codes. Even without any inappropriate upcoding, the incentive to generate additional risk codes reflect inefficiencies. The economically meaningful excess resource costs that go into generating these codes don’t create additional welfare. To the extent that a risk code generated from a review of charts is associated with less medical spending than a similar risk code that came about under the incentives of the FFS program, risk adjustment can end up being an inappropriately large transfer to private firms.

The trouble is that “fixing” risk adjustment is not easy. One solution would be to make risk adjustment a function of immutable characteristics such as age, race, sex, and geography. However, to the extent that there is still meaningful variability within these characteristics, firms would still have incentive to avoid sick individuals, conditional on these immutable characteristics (i.e., firms would still want to cream skim these immutable categories).

Another possibility is to move risk adjustment to a plan-level measure that is based on survey data. Such self-reported data from a random sample of enrollees would be harder to game than the existing system of risk codes. The challenge would be to identify the correct set of survey responses, but this is an area where policy should be focused.

6. Conclusion

If there is one thing that we have learned over the last several years, it is that health care is complicated. There are no easy ways to lower costs, increase access, improve quality, and encourage innovation. That said, the trade-offs inherent to these policy decisions don’t get any easier or less concrete by ignoring them. Efficient policy will only emerge from a careful consideration of these trade-offs.
References


Universal Basic Income (UBI) as a Policy Response to Current Challenges

AUTHORS
Melissa S. Kearney* and Magne Mogstad**

ABSTRACT
We briefly review the main motivations behind recent calls for a Universal Basic Income (UBI) in the United States and the main features of some current UBI proposals. We then argue that a UBI would be extremely expensive and yet do very little to reduce inequality or advance opportunity and social mobility. We argue that instead of a UBI, the federal government should pursue a pro-work strategy of income support that includes wage subsidies to low-wage workers, as well as cash and near-cash benefits targeted to the most needy individuals and households.

* University of Maryland: kearney@econ.umd.edu
** University of Chicago: magne.mogstad@gmail.com
1. Introduction

In recent years there has been renewed interest in the concept of a Universal Basic Income (UBI) provision. In its most basic form, a UBI is a guaranteed cash benefit that the government provides to all citizens. This is not a new idea, but one that has historically resurfaced from time to time. In the 1960s the economist Milton Friedman suggested and the Nixon administration considered a Negative Income Tax (NIT), which is a related policy idea that would give everyone some guaranteed level of income that would gradually be taxed away as own income increases. The longevity of such proposals can be attributed to the fact that various elements of UBI proposals appeal to both conservative and liberal political thinkers.

Three trends appear to be driving a renewed interest in UBI. First, some view a UBI as a reasonable response to growing inequality, to stem both economic and political unease. The American entrepreneur Andrew Yang claims the ratio of CEO pay to worker pay has risen from 20 to 1 in 1965 to 271 to 1 in 2016 (Yang, 2018). He notes that current levels of wealth and income inequality can be economically and politically destabilizing, and a UBI would provide a boost to the lowest-earners that could mitigate the effects of this inequality.

Second, some worry about the widespread elimination of well-paying jobs for many workers in the United States due to robots and other technological advancements. For this reason, the idea seems to have caught on among a number of tech futurist personalities. This also seems to be a main motivation behind the call for a UBI in the book *Give People Money* by American journalist Annie Lowrey (2018). She writes that it was not a question of “whether self-driving cars and other automated technology would start putting people out of work. It was when – and what would come next” (Lowrey, 2018). A UBI, Lowrey concludes, will provide a minimum standard of living for those shut out by automation and other forces, such as international trade.

A third, very distinct motivation for a UBI scheme is to streamline the current complicated and sometimes counterproductive system of U.S. transfer programs. American Enterprise Institute scholar Charles Murray is a prominent proponent of a UBI system for this reason. Murray proposes to convert all federal dollars currently spent on Social Security, Medicare, Medicaid, welfare, social services, and other programs into payments of $13,000 per year to every American aged 21 or older, which would be scaled back for higher earning individuals such that someone making $50,000 or more would receive a capped amount of $6,500 per year. In his 2016 book, Murray stipulates that under his proposal individuals would be required to use $3,000 of their UBI to purchase health insurance. He views the current system of programs as ineffective, and argues that instead, the government should simply give the money directly to people. Murray claims that savings from existing transfer program benefits and their administrative costs can be used to fund a UBI that provides for the poorest and in his view, is less likely to discourage work compared to some existing programs (Murray, 2016).
We view a UBI to be a suboptimal, and possibly harmful, policy response to all three of these challenges. A UBI in its most basic form would be massively expensive yet do little to reduce inequality or advance opportunity. Devoting that level of spending to targeted benefits, focusing on the poorest and those hardest hit by ongoing economic forces, and enacting polices dedicated to human capital development instead of mere redistribution would produce a much greater social return than a UBI.

First, on inequality and redistribution, a UBI is by design not ideal for redistributive purposes. For example, a UBI that paid $10,000 to every U.S. adult would cost about $2.5 trillion per year, well more than half the current federal annual budget. Furthermore, by giving money to everyone, there would be far fewer resources available to redistribute money and/or invest in the human capital of those with the most need. In calculations presented below, we show the practical trade-off between giving more money to a more targeted group of low-income individuals and less money to a more diffuse group including less needy individuals. We cite work by Hoynes and Rothstein (2019) documenting the loss of progressivity that would come from replacing our current system of transfer programs with a UBI. We further argue that programs that are universal do not accomplish as much in terms of generating social benefits as those that are targeted, citing evidence from examples such as childcare and early childhood education programs.

Second, on labor market trends and limited skills, the best long-term policy response is for the government to pursue a pro-work, pro-skills agenda, and devote resources to investing in the human capital development of children and economically disadvantaged groups of individuals. Such a policy emphasis would advance both individual economic security and aggregate productivity. However, it is critical that this long-term investment strategy be coupled with income support programs that provide wage subsidies for low-wage workers and limited cash and near-cash benefits for individuals who can’t work or who are temporarily out of work. This implies a pro-work agenda and targeted redistribution. UBI is neither.

Third, the safety net should not just be about redistribution, but also about investment in human capital and in the next generation. Programs should advance opportunity and economic mobility. Targeted in-kind programs and benefits have been shown to do that, especially when targeted to children. We are sympathetic to the argument that the existing safety net consists of a complicated array of different programs. To some extent, however, this complexity is a consequence of having different programs deliberately designed to serve different purposes and/or different needs. Even so, we are in favor of taking a holistic view of the panoply of safety net programs and reforming the entire system to work better in terms of both efficiency (namely, incentives) and equity (specifically, redistribution). Simplifying and improving the system does not, however, imply a UBI. It would be wholly counterproductive to address the complexity of the current system by replacing targeted programs that function as investments in human capital—thereby advancing the productivity of
future workers—with an income guarantee that would fail to advance opportunity and upward mobility. Instead, we should reform current programs to better achieve their desired goals.

In summary, for fiscal, efficiency, and equity reasons, the U.S. government should provide targeted benefits instead of universal benefits. And, it should not just provide cash, but rather invest in human capital and pursue redistribution through targeted spending on education, child care, health insurance, food vouchers, and housing assistance programs.

### 2. What Is Universal Basic Income (UBI)

A UBI true to its name would be unconditional and have no means test for eligibility. It would be given to every individual, regardless of their own or their family income. A related, but quite distinct, policy would be a conditional basic income program, or a Negative Income Tax (NIT), as it was named by University of Chicago economist Milton Friedman in his 1962 book *Capitalism and Freedom*. Under a conditional basic income scheme, the government would provide every individual a guaranteed income level, or a stipend, which would gradually be reduced as their earned income increased. This type of scheme was considered by the Nixon administration and evaluated in randomized controlled trial (RCT) social experiments in the 1960s and 1970s (discussed below).

We view a UBI to be related to an NIT, albeit much more expensive and with a very “leaky bucket,” to use the metaphor coined by Brookings scholar Arthur Okun (1975) to describe the “socio-economic leakages” of redistributive policy that arise from administrative costs, reduced savings or investment, or work disincentives. One of our main objections to a UBI is that by making the payment universal and unconditional, the government would be paying a lot of money to well off individuals, which would not serve any useful redistributive purpose but would divert public funds away from programs that could usefully advance human capital development and economic opportunity for truly needy Americans.

Here we outline key elements of some specific UBI proposals that have been put forward by UBI advocates in the past few years. As this discussion makes clear, many people who support UBI are actually arguing for a conditional basic income, not a universal one. We also offer some calculations showing how much it would cost the government to provide conditional basic income payments. These calculations highlight both the large expense of such schemes and the trade-off implicit in a conditional basic income scheme in terms of giving more money to fewer, lower-income people or less money to more people, including those with fairly high levels of income. For simplicity, the calculations assume no behavioral response. That
said, incorporating behavioral responses is likely to make the UBI proposals even more expensive, both directly through negative income effects on labor supply and indirectly through the distortionary costs of the additional taxes necessary to pay for the program.

Table 1 outlines key elements of six UBI proposals, a number of which actually provide important restrictions on the “universal” aspect of the UBI. These proposals differ in the size of the payment, as well as age and income eligibility requirements. For instance, a common version offered by the labor leader Andy Stern, journalist Annie Lowrey, and entrepreneur Andrew Yang would send a $1,000 check to every adult (or deposit the money in their personal account) every month. By not phasing out benefits with earned income, these programs trade off targeting efficiency with productive efficiency: More people get the benefit (making it more expensive and less targeted on the needy), but the program only discourages work insofar as income discourages work. The program does not additionally discourage marginal work by taxing away benefits as earnings increase. Yang proposes funding a UBI by consolidating some welfare programs and implementing a Value-Added Tax (VAT) of 10%. He would allow current welfare and social program beneficiaries to be given a choice between their current benefits or $1,000 cash unconditionally. Even these simple plans violate the universal aspect of the UBI, strictly speaking. Philosophers Philippe Van Parijs and Yannick Vanderborght offer one of the only truly universal basic income proposals, in which all citizens, regardless of age, income, and working status receive monthly installments of an annual income equal to one-quarter of the country’s GDP per capita (roughly $15,000 per year in the United States in 2017).

Proposals by Charles Murray and entrepreneur Chris Hughes include further restrictions on eligibility and thus more closely resemble an NIT than a UBI. Under Murray’s plan, individuals would receive an annual transfer of $10,000 ($13,000 inclusive of the requirement to pay $3,000 for health insurance), paid out in monthly payments, until that person’s annual income exceeds $30,000. Each additional dollar earned after that threshold reduces the payment by 30 cents, although all who make more than $60,000 receive the minimum of $6,500. Hughes’ proposal is more restrictive, offering a monthly payment of $500 to working adults in households that make less than $50,000 per year, with a few exceptions for adults with young children, older dependents, or those enrolled in college.
Table 1: Proposals for a Universal Basic Income or Conditional Basic Income

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Transfer</td>
<td>$10,000</td>
<td>$12,000</td>
<td>$12,000</td>
<td>$12,000</td>
<td>$6,000</td>
<td>Approx. $15,000*</td>
</tr>
<tr>
<td>Phase-out begins</td>
<td>$30,000</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>$50,000</td>
<td>n/a</td>
</tr>
<tr>
<td>Phase-out rate</td>
<td>11%*</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>100%</td>
<td>n/a</td>
</tr>
<tr>
<td>Age restrictions</td>
<td>21+</td>
<td>18+</td>
<td>18+</td>
<td>18+</td>
<td>18+</td>
<td>“All citizens”</td>
</tr>
<tr>
<td>Additional notes</td>
<td>*Phase out UBI to $6,500</td>
<td>Restricted to “working adults”</td>
<td>*Transfer set to one-quarter of GDP per capita</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 provides key elements of five pilot studies that are currently under way in various locations. These pilot programs are being conducted to test the feasibility and effectiveness of unconditional cash transfers in advanced economies. There are two ongoing experiments in the United States. In February 2019, a study funded by the Economic Security Project, co-chaired by Chris Hughes, began distributing $500 per month to around 100 randomly selected program applicants in Stockton, California (Yoon-Hendricks, 2019). Eligible recipients needed only to be adults living in one of Stockton’s lower-income Census tracts. The income disbursements are scheduled to last 18 months. A much larger initiative, directed by Silicon Valley start-up incubator Y Combinator, plans to provide $1,000 per month to 1,000 adults aged 21-40 across two states, lasting either 3 or 5 years (YC Research, 2018). Importantly, planners of this study aim to obtain waivers from welfare administrators to exclude this income from determining eligibility in targeted transfer programs. As Hoynes and Rothstein (2018) note, given that most UBI plans are partially funded through the elimination of other transfer programs, these results will reflect the effects of an income supplement rather than a true UBI.

Internationally, two government-run pilot programs have concluded, both ending earlier than initially planned due to political or financial difficulties. Beginning in early 2017, Finland offered monthly checks of 560 euros (around $650) to 2,000 randomly selected, unemployed persons between the ages of 25 and 58. However, the government opted to end the program at the end of 2018, a reflection, as New York Times reporter Peter Goodman put it, “of public discomfort with the idea of dispensing government largess free of requirements that its recipients seek work” (2018). Researchers released initial findings that recipients of this income were less likely to be employed, but did self-report higher levels of psychological well-being, although researchers did note concerns about low survey response rates among recipients (Bershidsky, 2019). Furthermore, this program also served as a supplement to Finland’s welfare benefits, such that recipients who were eligible for additional unemployment benefits received them along with the cash supplement. Ontario’s plan to test a UBI among 4,000 low-income participants was also cancelled 2 years early, at the end of 2018 (Frazee, 2018).
Table 2: Completed, Ongoing, and Planned UBI Pilot Programs

<table>
<thead>
<tr>
<th></th>
<th>Finland</th>
<th>Stockton, CA</th>
<th>Ontario</th>
<th>Switzerland</th>
<th>2 U.S. States (Y Combinator)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Annual Transfer</strong></td>
<td>6,720€ ($7620)</td>
<td>$6,000</td>
<td>$16,989 CAN individuals $24,027 CAN couples</td>
<td>26,280€</td>
<td>$12,000</td>
</tr>
<tr>
<td><strong>Phase-out begins</strong></td>
<td>n/a</td>
<td>n/a</td>
<td>$0</td>
<td>n/a</td>
<td>County median income</td>
</tr>
<tr>
<td><strong>Phase-out rate</strong></td>
<td>n/a</td>
<td>n/a</td>
<td>50%</td>
<td>n/a</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Age restrictions</strong></td>
<td>25-58</td>
<td>n/a</td>
<td>18+</td>
<td>n/a</td>
<td>21-40</td>
</tr>
<tr>
<td><strong>Treatment group size</strong></td>
<td>2,000 individuals</td>
<td>100 families</td>
<td>2,000 individuals</td>
<td>TBD</td>
<td>1,000 individuals</td>
</tr>
<tr>
<td><strong>Dates</strong></td>
<td>2017-2018</td>
<td>2019-2020</td>
<td>2017-2018</td>
<td>TBD</td>
<td>2020-2023 or 2025</td>
</tr>
<tr>
<td><strong>Interaction with Welfare Payments</strong></td>
<td>Basic income deducted from transfer payments</td>
<td>UBI supplements transfer income</td>
<td>Replaces most transfer programs</td>
<td>Basic income deducted from transfer payments</td>
<td>Seeking waivers for UBI to supplement transfer income</td>
</tr>
<tr>
<td><strong>Additional Notes</strong></td>
<td>Government declined to extend trial in 2018</td>
<td>Trial ended 2 years early</td>
<td>Smaller transfers for younger age brackets</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. UBI Is Not an Ideal Tool for Redistributive Goals

UBI is by design not ideal for redistributive purposes. Given that resources are necessarily limited and the government would have to operate within a UBI budget, the more that is given universally, the less there is to give to the truly needy. But even before we dissect that point, we note that the cost of any UBI program that would make a material difference for household income would be massive. For example, a UBI that paid $10,000 to every person (adults and children alike) in the United States would pay over $3 trillion in benefits per year, or more than three-quarters of the current federal annual budget. This is more than the sum of costs for all current federal income support programs plus Medicare and Medicaid. Thus, a UBI of this magnitude could not be paid for by replacing all existing social safety net programs.

Using the 2018 Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC), we calculate annual payments across four hypothetical UBI-style programs: (A) $10,000 to all adults age 18 or older; (B) $10,000 to all adults with earned income of less than $10,000 and then phased out at 30%; (C) $10,000 to all adults with earned income of less than $20,000 and then phased out at 30%; (D) $10,000 to all adults with earned income of less than $20,000 for individuals or family earned income of less than $40,000 for marrieds and then phased out at 30%.
The total payment costs for each plan are presented in Table 3, with rows reporting the amount of payments that would be given to people in each income quintile based on family income. Figure 1 plots the share of transfer payments distributed to each quintile of family income for these four hypothetical UBI-style programs. For comparison, we include in the table and figure estimates of existing transfer program income, as reported in the CPS. Plan A—a true universal basic income with no phase out—is obviously the most expensive, coming in at $2.49 trillion and the least progressive, as there is no targeted or conditional element. Plan B, which would give $10,000 to all adults with earned income under $10,000 and then be phased out at a 30% benefit reduction rate, is much more targeted than the universal plan and consequently, much less costly, though still $1.4 trillion. About 40% of benefit payments would go to individuals in the lowest income quintile and less than 10% would go to individuals in the highest income quintile. Plan C is less targeted, paying $10,000 annually to individuals with less than $20,000 in earned income, and phased out at a 30% benefit reduction rate. That leads to higher costs of $1.6 trillion, coming from higher payments to individuals closer to the middle and upper end of the income distribution. Plan D has separate thresholds for single and married individuals, which is more in line with the way most current transfer program eligibility and benefit amounts are calculated based on family income. This plan is roughly comparable in payment amounts and distribution to Plan B, but it allocates less transfer income to higher-income families and is less costly overall.

Table 3: Cost of Payments for Four Hypothetical UBI-Style Programs, in Trillions of Dollars

<table>
<thead>
<tr>
<th>Family Income Quintile</th>
<th>Plan A</th>
<th>Plan B</th>
<th>Plan C</th>
<th>Plan D</th>
<th>Existing Transfers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. less than 13,520</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
<td>0.60</td>
</tr>
<tr>
<td>2. 13,520-43,330</td>
<td>0.48</td>
<td>0.33</td>
<td>0.41</td>
<td>0.43</td>
<td>0.19</td>
</tr>
<tr>
<td>3. 43,330-77,401</td>
<td>0.48</td>
<td>0.22</td>
<td>0.28</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>4. 77,401-130,096</td>
<td>0.49</td>
<td>0.16</td>
<td>0.21</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>5. greater than 130,096</td>
<td>0.49</td>
<td>0.12</td>
<td>0.14</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Total ($trillions)</strong></td>
<td><strong>2.49</strong></td>
<td><strong>1.38</strong></td>
<td><strong>1.60</strong></td>
<td><strong>1.20</strong></td>
<td><strong>1.00</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Share of GDP ($20.5T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.15</td>
</tr>
</tbody>
</table>

1 Estimated transfer payments reported in this table come from reported benefits in the CPS received through Supplemental Security Income (SSI), Social Security Old Age and Survivors (SSOAS), Social Security Disability Insurance (SSDI), cash welfare, food stamps, and housing assistance, as well as the estimated EITC payment calculated in the census tax module. Estimated transfer payments using the CPS data total $1.004 trillion, which is lower than the figure based on administrative data reported in Table 4. This discrepancy is likely due to the under-reporting of benefits in survey data, as documented by Meyer, Mok, and Sullivan (2009).
Table 4 reports annual expenditures on existing transfer programs. The Earned Income Tax Credit (EITC) disbursed $67 billion in 2016, from the most recent estimates. The Supplementary Security Income Program (SSI) transferred roughly $55 billion in 2018, primarily to disabled adults, with smaller sums to elderly individuals and disabled children, while Social Security (SS) expenditures totaled over $940 billion that same year. Temporary Assistance for Needy Families (TANF), Supplemental Nutritional Assistance Program (SNAP), and Section 8 and Public Housing Assistance spending on benefit payments reached $7.1, $60.4, and $27.7 billion in 2018, respectively.

Figure 1: Distribution of Transfer Payments Under Four Hypothetical UBI-Style Programs and Current Set of Programs, by Family Income Quintile

Table 4 reports annual expenditures on existing transfer programs. The Earned Income Tax Credit (EITC) disbursed $67 billion in 2016, from the most recent estimates. The Supplementary Security Income Program (SSI) transferred roughly $55 billion in 2018, primarily to disabled adults, with smaller sums to elderly individuals and disabled children, while Social Security (SS) expenditures totaled over $940 billion that same year. Temporary Assistance for Needy Families (TANF), Supplemental Nutritional Assistance Program (SNAP), and Section 8 and Public Housing Assistance spending on benefit payments reached $7.1, $60.4, and $27.7 billion in 2018, respectively.
Table 4: Annual Expenditures on Existing Transfer Programs

<table>
<thead>
<tr>
<th>Program</th>
<th>Expenditures ($billions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earned Income Tax Credit (EITC)</td>
<td>66.7</td>
</tr>
<tr>
<td>Supplemental Security Income Program (SSI)</td>
<td></td>
</tr>
<tr>
<td>SSI-Aged</td>
<td>5.5</td>
</tr>
<tr>
<td>SSI-Children</td>
<td>9.4</td>
</tr>
<tr>
<td>SSI-Disability</td>
<td>39.6</td>
</tr>
<tr>
<td>Social Security</td>
<td></td>
</tr>
<tr>
<td>SS-Old Age and Survivors (SSOAS)</td>
<td>798.7</td>
</tr>
<tr>
<td>SS-Disability Insurance (SSDI)</td>
<td>142.8</td>
</tr>
<tr>
<td>Temporary Assistance for Needy Families (TANF)</td>
<td>7.1</td>
</tr>
<tr>
<td>Supplemental Nutritional Assistance Program (SNAP)</td>
<td>60.4</td>
</tr>
<tr>
<td>Section 8 and Public Housing Assistance</td>
<td>27.7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,157.9</strong></td>
</tr>
<tr>
<td><strong>Total without SS-Old Age and Survivors</strong></td>
<td><strong>359.2</strong></td>
</tr>
</tbody>
</table>


Many existing programs in the United States have categorical eligibility requirements. This means that characteristics of individuals or families are used to determine “need,” beyond just income. The theoretical work by Nobel Prize-winning economist George Akerlof demonstrated that using markers of need beyond income allows for enhanced efficiency because the government can then more readily avoid giving transfer payments to individuals who are able to work but choose not to. Using what he referred to as “tagging” mechanisms helps the government distinguish between those who need assistance and those who instead choose not to exert effort. The presence of children in a family and a medically verified disability condition are two such tags that are currently used to identify individuals and families in need. Using language more common to political discourse than theoretical considerations of program design, we might refer to these classes of individuals as the “deserving poor.” A UBI would move away from this type of categorization to paying income benefits universally or based solely on income. Some would consider this a “pro” because then the government wouldn’t be in the business of trying to determine who was “deserving,” others would consider it a “con” because it would give transfer payments to people who choose not to work, rather than having real underlying need.
Moving from our current system of income support programs to a UBI would mean shifting existing transfer payments away from low-earners to both non-earners and higher earners, as well as away from families with children, the disabled, and elderly to able-bodied individuals without children (who currently receive very little by way of income support). Hoynes and Rothstein (2019) compare the distributional effects of a UBI and the current system of U.S. income support programs. They consider the existing safety net as consisting of means-tested welfare programs (Temporary Assistance to Needy Families (TANF) and Supplemental Nutritional Assistance Program (SNAP)); disability programs (Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI)); Social Security Old Age and Survivors (SSOAS); and in-work tax credits (Earned Income Tax Credit (EITC) and Child Tax Credit (CTC)). They exclude all in-kind programs other than SNAP, most notably public health insurance and also housing support. They document that the current social safety net gives higher transfers to the elderly and disabled, higher transfers to those with children compared to those without, and higher transfers to those with low earnings. Consequently, replacing current income support programs with a UBI would result in a relative redistribution from low-earners to zero earners, but, as they write, the “first-order effects would be a massive redistribution up the earnings distribution, along with a redistribution from the elderly and disabled towards those who are neither, primarily but not exclusively those without children.”

Another scenario for converting current programs to a UBI would be to preserve Social Security, Medicare, and Medicaid, and instead eliminate only the EITC, CTC, TANF, SNAP, disability insurance (SSDI), and unemployment insurance (UI). As Hoynes and Rothstein (2019) point out, these six programs combined would only cover about one-fifth of the cost of a UBI. The remainder would need to be funded through cuts to non-transfer government expenditures or through tax increases. If instead, we were to implement a non-universal conditional basic income program and eliminate just the six programs—EITC, CTC, TANF, SNAP, SSDI, and UI—the budget implications could be minimal, but there would be a tremendous loss of existing transfers to disabled individuals and families with children.

Robert Greenstein of the Center for Budget and Policy Priorities (CBPP) has also made this point. He argues that political passage of UBI would require scrapping existing safety net programs such that the likely result would be to increase poverty. The CBPP calculates that a program that eliminates all means-tested transfer programs outside of health care would only fund $1,582 per person, well below the income needed to keep a family out of poverty. Greenstein instead proposes to shore up income support for the most needy individuals and families with an expansion of existing targeted programs. He rejects the claim that is sometimes made that a universal program would be more politically feasible than expanding means-tested transfer programs, noting that restrictions on Social Security benefits have been passed in recent decades (for example, increasing the retirement age), while programs such as the EITC have been expanded. He writes that “instead of aiming for an expensive
universal income provision that may end up transferring money from the poor to wealthier families, we should take advantage of the relatively low cost of expanding targeted programs—and their proven record of lifting families out of poverty—as a method of raising and broadening existing income floors” (Greenstein, 2019).

### 4. UBI Is Not the Ideal Response to Employment and Wage Trends

There is no denying that economic trends over the past few decades have disproportionately benefited highly educated, high skill workers and that some distinct groups of Americans, most notably those without a college level of education, have seen their wages and employment prospects weakened. Among high school dropouts, high school graduates, and those with some college but no degree, real weekly earnings among full-time, male workers in 2018 were 10 to 20 log points below their real levels in 1980, while those with a bachelor degree experienced wage increases of about 10 log points (Autor, 2019). In terms of job polarization, there has been an overall shift from middle-wage jobs to higher-wage jobs. But, as David Autor has shown, the shift upward from middle to high paying jobs is driven by workers with college degrees. Among workers without a college degree, almost all of the decline in middle-wage employment reflects a shift to lower paid work. These wage trends reflect both market forces, including technological advancements and globalization, as well as institutional changes, including a decline in union representation and a rise in outsourcing.

Employment rates among prime-age men and women have been falling steadily for decades, notwithstanding the growth in employment through the recent cyclical recovery. Between 1999 and 2015, for each 10-year age group between the ages of 25 and 54, the employment rate fell 3 to 4 percentage points (Abraham and Kearney, forthcoming). Employment rates have fallen the most among men and those with a high school degree but no college degree. For instance, employment rates among men aged 25 to 34 with a high school degree fell by more than nine percentage points. These declines in employment have led many observers to question what types of policies might be implemented to increase labor force participation rates. A UBI or UBI-style proposal to give unconditional cash would achieve exactly the opposite effect, since both theory and empirical evidence demonstrate that giving unconditional cash will lead to lower rates of work, not higher.

A UBI or UBI-style proposal in response to these wage and employment trends would do nothing to address their underlying causes. Rather, it would be a band-aid solution of giving cash instead of enhancing skills or increasing bargaining power. Furthermore, a UBI’s contrast with a wage subsidy, such as an Earned Income Tax Credit (EITC), is striking. A large body of evidence shows that the current EITC encourages labor force participation, especially among single women. Wage subsidies—either
through the EITC or through a payroll subsidy like the kind proposed by Furman and Swagel (2018)—would increase work effort by raising the take-home pay of low-wage workers. A program of wage subsidies has better efficiency properties and equity properties than a UBI or even a conditional basic income program.

Some proponents argue that an award of unconditional cash will not necessarily reduce labor supply. One reason is because individuals could use the cash income from a UBI to enroll in school or engage in other skill enhancing activities to increase their future labor supply. Another more commonly offered defense appeals to research suggesting that the reduction in work owing to an increase in income tends to be small. But, the most commonly cited research making this point is either based on decades-old data or from small guaranteed basic income and transfer programs, such as payouts from U.S. state or Native American dividend programs. Hum and Simpson (1993) reviewed evidence from U.S. and Canadian NIT experiments from the 1960s and 1970s, finding very small, negative effects of these income maintenance programs on labor supply. Given both the continued polarization of the labor market and the significant evolution of attitudes about work since the 1970s, discussed below, it is not clear that these parameter estimates are applicable to contemporary debates about labor market effects. It is also quite possible that the negative effects of a guaranteed income on labor supply, especially among less-educated men, would be larger today than they were 40 to 50 years ago.

There are also a number of studies examining the labor supply effects of public dividend payment programs, namely the Alaska Permanent Fund Dividend and the Eastern Band of Cherokee’s casinos dividend payout. Alaska’s program provides on average $1,000-$2,000 to all residents of the state, while the Cherokee’s program average $4,000 per year. Researchers find small labor market effects of these transfers as well, with insignificant changes in labor force participation among recipients of the casino payouts and only small employment declines among some sets of workers in Alaska (Marinescu, 2017; Jones & Marinescu, 2018). Although universally available, these transfers could not support even a modest standard of living—indeed, a shift into part-time work might be the only reasonable labor market adjustment from such a program—so these estimates also appear ill-equipped to address current proposals.

Studies of transfers that are more comparable in size to the types of UBI payments being proposed imply more negative labor supply effects. For example, a study of lottery winners (Imbens, Rubin, & Sacerdote, 1999) find that, with an average annual prize of $26,000, each $100 in additional earnings reduced labor market earning by $11. A more recent study of lottery winners in Sweden also provides evidence of reduced earnings in response to winning a lottery prize. This study finds that winning a lottery prize leads to an immediate and persistent reduction in earnings (Cesarini, Lindqvist, Notowidigdo, & Östling, 2017). In addition, the effects of any guaranteed-income program are likely to most strongly affect those marginally attached to the
labor force. On this point, the lessons from expanded access to disability insurance payments is potentially instructive. Economists have found that the marginal beneficiary of a disability insurance award would have been almost 30 percentage points more likely to work had they not received benefits (for example, Maestas, Mullen, & Strand, 2013).

Not all UBI proponents try to argue that a UBI would not lead to a reduction in work effort. Some UBI proponents, including Albert Wenger, a UBI advocate and venture capitalist, explicitly promote the concept of a UBI on the grounds that it would liberate people from the need to work. This is a policy position with which we fundamentally differ in terms of the goals itself. We do not view it as a goal of public policy or government spending to subsidize the able-bodied who simply choose not to work. Other UBI proponents take a less radical view, aiming not to liberate people from work itself but instead from lousy jobs. But if that’s the goal, then why not support wage subsidies and other regulations on workplace conditions? Such an approach would improve the work experience and increase take-home pay without discouraging work.

5. Universal Programs Are Ineffective in Advancing Opportunity and Social Mobility

The social safety net should not be just about redistribution, but also about improving human capital and fostering the skills of the next generation. Targeted cash and in-kind transfer programs have been shown to do that. A growing body of evidence suggests that economic security programs can blunt the negative effects of poverty and bring poor children closer to equal opportunity. Numerous studies document the positive effects of EITC payments on a range of outcomes, including children’s academic performance, infant health, and maternal mental health. This body of evidence is reviewed in Nichols and Rothstein (2015). Crucially, these are income payments given to low-income households. We are not aware of any evidence showing that incremental income payments paid to higher income people similarly produces positive social returns. Studies of the Medicaid public health insurance program and the national food stamps program document positive long-term effects both for the recipients and their children (c.f., Brown, Kowalski, & Lurie, 2015; Boudreaux, Golberstein, & McAlpine, 2016; Hoynes, Schanzenbach, & Almond, 2016).

Furthermore, safety net programs tend to have the highest social returns when they are targeted to children from disadvantaged families. For example, a large body of evidence supports a public subsidy of high-quality, childcare programs targeted to disadvantaged populations. At current quality levels and costs, their social benefits greatly exceed their social costs (Ludwig & Phillips, 2007). Some proponents of universal child care have used such estimates to make the case for expansion. But,
using the evidence from targeted programs for the promotion of universal ones is problematic for a number of reasons. First, the widely cited, targeted programs generally offer quality levels that are not typically found in programs that are offered to all children. Therefore, it is not clear that universal childcare can deliver similar benefits to disadvantaged children. Second, even if universal programs could offer similar levels of quality in terms of care and education, it may be that the benefits for middle- or upper-class children do not exceed the costs.

Indeed, a number of studies show that children from middle- or upper-class families benefit little if any from subsidized childcare. An important reason is that children in these families are already receiving high quality childcare in the absence of government subsidies, and, thus, public subsidies do little to alter parental behavior or increase investments in children. A study by Havnes and Mogstad (2015) shows that when childcare subsidies are provided universally in Norway, the observed benefits are positive for children of low- and middle-income families and negative for children of higher-income families. Similarly, Cascio and Schanzenbach (2014) have made the point that expanding early childhood education programs to low-income children who currently do not have access to such programs will lead to improved early childhood experiences, but for higher-income children who are already likely to be participating in private, early childhood programs, the effect will likely be non-existent, or potentially even negative if children are moved to lower quality/lower cost programs. Similar arguments can be made for targeted programs other than subsidized child care and early childhood education.

6. Conclusion

Advocates of a UBI make three broad arguments to promote the program. First, it is argued to be a reasonable response to growing inequality, to stem both economic and political unease. Second, it is supposed to ensure a minimum standard of living for those shut out by automation and other disruptive forces, such as international trade. A third, very distinct motivation for a UBI scheme is to streamline the current complicated and sometimes counterproductive system of U.S. transfer programs.

As argued above, a UBI is almost surely a sub-optimal, and likely harmful, policy response to all three of these challenges. A UBI in its most basic form would be massively expensive yet do little to reduce inequality or advance opportunity. Without major cost savings, U.S. federal tax revenue would have to be increased radically, which would impose large distortionary costs on the economy. Sacrificing all other social programs for the sake of a UBI is also a poor idea. Such programs exist to address specific problems, such as the vulnerability of the elderly, children, and people with disabilities. Focusing spending on targeted benefits and polices dedicated to human capital development instead of merely on redistribution is likely to produce a much greater social return than a UBI.
References


Wealth Taxation: An Overview of the Issues

AUTHOR

Alan D. Viard

ABSTRACT

Two Democratic presidential candidates, Senator Elizabeth Warren (D-Massachusetts) and Senator Bernie Sanders (I-Vermont), have proposed annual wealth taxes on extremely wealthy households. Annual wealth taxes have been adopted in a number of European countries (many of which later repealed them), but not in the United States. Although the proposed wealth tax rates appear low, they are equivalent to high-rate income taxes. Due to the pronounced concentration of wealth in the United States, a wealth tax would be highly progressive. The tax would probably reduce national saving and investment to some extent, although capital inflows from abroad would ameliorate the investment reduction. Congress would likely add exemptions for selected assets, which would be distortionary and diminish the tax’s revenue yield. The tax would face compliance and administration challenges due to undervaluation and concealment of assets and it might be ruled unconstitutional in the absence of suitable modifications. Although those challenges would probably not be insurmountable, it would be simpler and more prudent to pursue any desired increase in tax progressivity through reforms of the income tax and estate and gift taxes.

* American Enterprise Institute: aviard@aei.org. I am grateful to Amy Ganz, Melissa Kearney, Erin Melly, Harvey Rosen, and Victor Thuronyi for helpful comments.
1. Overview of Wealth Taxes

This chapter examines proposals to impose annual taxes on wealth or net worth. Under a wealth tax, households would pay tax each year based on their net worth, which is the fair market value of their assets minus the fair market value of their liabilities. Wealth taxes generally would apply only to wealth above an exemption amount.

Annual wealth taxes have not been used in the United States. However, the federal government and many states impose estate and gift taxes, which are essentially once-per-lifetime wealth taxes that are paid when wealth is transferred through gift or bequest.

Many European countries have adopted annual wealth taxes, although a majority of those countries subsequently repealed them. The Organisation for Economic Co-operation and Development (OECD) reported in 2018 that only four of its member countries—France, Norway, Spain, and Switzerland—were imposing annual wealth taxes in 2017. As Bunn (2019) observed, however, six OECD countries actually had wealth taxes, as the Netherlands imposed a wealth tax embedded within its income tax system and Italy imposed a tax on assets that Italians held abroad. In 2018, France repealed its wealth tax and Belgium introduced one, leaving the number of OECD countries with wealth taxes unchanged at six. The other OECD countries that have repealed wealth taxes are Austria, Denmark, Finland, Germany, Iceland, Ireland, Luxembourg, and Sweden. As discussed below, the repeals were generally motivated by administration and compliance difficulties, undesired behavioral responses such as emigration, and disappointing revenue yields.

In January 2019, Senator Elizabeth Warren (D-Massachusetts) proposed an annual wealth tax as part of her campaign for the 2020 Democratic presidential nomination. Her proposal (Warren, 2019a) featured a tax rate of 2% per year on wealth in excess of a $50 million exemption amount, with a rate of 3% per year on wealth in excess of $1 billion. In September 2019, Senator Bernie Sanders (I-Vermont), who is also seeking the Democratic presidential nomination, proposed a wealth tax on households with wealth above $16 million for singles ($32 million for married couples). The tax rates in his proposal (Sanders, 2019) would start at 1% per year and would reach 8% per year on wealth above $5 billion ($10 billion for married couples). In November 2019, Senator Warren revised her proposal, raising the tax rate on wealth in excess of $1 billion to 6% per year (Warren, 2019b). Congressional Democrats have generally been reluctant to embrace the proposed wealth taxes (Elis, 2019).

---

1 This chapter does not examine proposals to impose one-time wealth taxes, sometimes referred to as capital levies.

2 State and local governments impose annual property taxes, but they are significantly different from wealth taxes. They primarily apply to real property and they do not allow a deduction for the taxpayer’s liabilities. Pomerleau (2019) examined the differences between property taxes and wealth taxes.
Multiple public opinion polls have found majority support for wealth taxation, with the strongest support from political liberals and Democrats. A Business Insider poll taken shortly after the Warren proposal was released found that it was supported by 54% of the public and opposed by 19%. The proposal was supported by 76% of self-proclaimed liberals, 36% of self-proclaimed conservatives, and 56% of those who did not identify with either category (Bryan, 2019). In April, a Quinnipiac University Poll (2019) found support from 60% of the public, including 82% of Democrats, 63% of independents, and 32% of Republicans. Sarin and Summers (2019b) noted, however, that the estate and gift taxes lost political support when they came under sustained political attack and suggested that support for wealth taxes may also erode as they receive more scrutiny.

Wealth taxation has drawn support even among affluent households. A June 2019 CNBC survey of persons with net worth above $1 million found that 60% supported the Warren proposal, including 88% of Democrats, 62% of independents, and 36% of Republicans (Frank, 2019). Of course, many of those persons would not be subject to the tax because their wealth is below the Warren proposal’s $50 million exemption amount. In June 2019, however, 19 multimillionaires and billionaires who would pay substantial taxes under the Warren proposal released a letter supporting the proposal (Bowditch et al., 2019). In September 2019, Microsoft Corp. founder Bill Gates said that he “would not be against” a wealth tax (Metcalf & Schatzker, 2019).

The proposed wealth taxes would apply to the worldwide wealth of U.S. citizens (even if living abroad) and of non-citizens who have U.S. permanent resident status or spend specified amounts of time in the United States. The Warren proposal would allow tax payments to be deferred for 5 years, with interest, to address the (perhaps unlikely) possibility that some taxpayers might lack sufficient liquidity to immediately pay the tax. The proposals also include enforcement and anti-avoidance provisions, as discussed below.

2. Effects of Wealth Taxation

2.1 Interpreting Wealth Tax Rates

Although wealth tax rates of 6% or 8% may appear to be low, that appearance is deceiving. It is important to realize that the rates are actually 6% or 8% per year. Because a flow of taxes is imposed on a stock of wealth, the tax rate cannot be stated without specifying a time unit. For a household with constant wealth, under a 6% annual wealth tax, tax equal to 6% of wealth would be paid over the first year, but a cumulative tax equal to 60% of wealth would be paid over a decade. In contrast, no time unit is needed to state income tax rates because a flow of taxes is imposed on a flow of income. Under a 30% income tax, tax equal to 30% of each year’s income would be paid each year and tax equal to 30% of each decade’s income would be paid each decade.
A useful way to interpret wealth tax rates is to translate them into equivalent income tax rates. For a taxpayer who holds a long-term bond with a fixed interest rate of 3% per year, a 6% per year wealth tax is equivalent to a 200% income tax because the tax equals 200% of the taxpayer’s interest income. Similarly, an 8% per year wealth tax is equivalent to a 267% income tax.

The tax-rate translation is more complicated for risky investments. Suppose that, alongside her holdings of the 3% bond, the taxpayer holds a stock with an annual return that could fall anywhere between 2% and 10%, with an expected value of 6%. The 6% per year wealth tax could end up being anywhere from 60% to 300% of the stock’s return. It is not immediately clear what income tax rate the taxpayer would perceive as equivalent to the wealth tax in advance, when the stock return is uncertain. At first glance, it may seem that the 6% per year wealth tax is equivalent to a 100% income tax rate, which would also result in an expected tax payment of 6% per year. As Sarin and Summers (2019b) observed, however, the wealth tax payment, which is a fixed fraction of wealth, cannot be compared to the expected value of the uncertain income tax payment, because the payments have different risk characteristics.

The puzzle can be solved by observing that the taxpayer should normally hold a mix of the stock and the bond that makes her equally content with both assets on the margin. In that case, the risky stock return and the safe bond return must be equally attractive on the margin, despite the differences in their expected returns. A 200% tax on the bond return (which is equivalent to the 6% per year wealth tax) and a 200% tax on the risky stock return must then be equally burdensome. The wealth tax’s equivalent income tax rate is therefore 200% for the stock as well as for the bond. Similarly, an 8% per year wealth tax would be equivalent to a 267% income tax for the stock as well as for the bond. The conclusion that wealth taxes’ equivalent income tax rates should be computed by treating all assets as earning the same return as safe assets must be modified if financial frictions or other factors prevent investors from choosing portfolios that make them equally content with all assets on the margin. However, the conclusion is likely to be a reasonable approximation, particularly for wealthy taxpayers, who are likely to face few financial frictions. Bulow and Summers (1984) provide further discussion of the taxation of risky returns.

Wealth taxes can be equivalent to extremely high income tax rates. Moreover, the wealth tax would be imposed in addition to the income tax, making total tax rates even higher. Whether or not such high rates are viewed as desirable, it is important to understand them.

The proposed top tax rates in Warren (2019b) and Sanders (2019) are far higher than European wealth tax rates. Bunn (2019) reported rates of 0.15% per year in Belgium, 0.2% to 0.76% per year in Italy, 0.61% to 1.61% per year in the Netherlands, 0.85% per year in Norway, and 0.2% to 2.5% per year in Spain.
The fact that wealth tax payments, unlike income tax payments, would be the same fraction of wealth for investors with high returns and those with low returns has several implications. The failure to impose additional tax on investors who earn higher returns would make the tax less effective at its goal, discussed below, of curbing wealth concentration. Kaeding and Pomerleau (2019) criticized the wealth tax for not imposing additional tax on investors who, due to monopoly power or special skills, can command windfall returns beyond the returns needed to maintain investment incentives, arguing that such windfall returns can often be taxed with little economic harm. However, Saez and Zucman (2019b) pointed out that some of the apparent windfall returns may be a payoff to past entrepreneurial activity and that the wealth tax’s failure to impose higher tax on such returns helps maintain incentives for such activity. Guvenen, Kambourov, Kuruscu, Ocampo-Diaz, and Chen (2019) argued that the wealth tax’s failure to impose higher tax on entrepreneurs with special skills is desirable because it reallocates funds toward those entrepreneurs, thereby increasing aggregate productivity.

2.2 Progressivity and Wealth Concentration

An annual wealth tax would be highly progressive. A broad range of estimates find that the U.S. wealth distribution is extremely concentrated and has become more concentrated in recent decades, although there is disagreement about the magnitude. Saez and Zucman (2016) estimated that the wealthiest 1% of households owned 42% of national wealth in 2012, with the top 0.1% owning 22% and the top 0.01% (the top one ten-thousandth) owning 11%. Smith, Zidar, and Zwick (2019, 26) found somewhat less concentration, estimating that the top 1% owned 31% of national wealth in 2014, with the top 0.1% owning 15% and the top 0.01% owning 7%.

Saez and Zucman (2019a) estimated that Senator Warren’s proposed tax would apply to 75,000 households, approximately 0.06% of all households. They estimated that those households own 10% of national wealth. Saez and Zucman (2019d) estimated that Senator Sanders’ proposed tax would apply to 180,000 households, approximately 0.15% of all households. Even if a small portion of the wealth tax burden were shifted to workers in the form of lower wages (a possibility discussed below), the tax would remain highly progressive.

Progressivity may be desired because it allows taxes to be collected from those who can best afford to pay them. Economists generally assume that the loss of utility, or well-being, from a dollar tax payment is smaller for persons with more economic resources. Holding everything else equal, raising revenue from a small group of top wealth holders would therefore involve less loss of well-being than raising the same revenue from a broader group. Similarly, if the revenue raised from a small group of wealth holders were transferred to a broader group, the recipients’ gain in well-being would exceed the wealth holders’ loss of well-being, unless the transfer caused large inefficiencies. Collecting additional revenue from top wealth holders might also be considered a move toward tax fairness.
Many supporters of wealth taxation advance a different rationale, arguing that the reduction of wealth concentration is inherently beneficial. Senator Sanders invoked that rationale on the day that he released his plan, tweeting “There should be no billionaires” (Cillizza, 2019). One justification for objecting to wealth concentration is that it places too much political power in the hands of a small group. However, this argument does not provide a convincing rationale for wealth taxation.

To begin, the rationale can be challenged on normative grounds. It is far from clear that the government should define the proper distribution of political power in a free society. One might ask whether the government should seek to weaken other groups, such as the media, universities, and think tanks, which are also likely to have power disproportionate to their numbers. In any event, a wealth tax is unlikely to have a significant impact on the distribution of political power. As Sarin and Summers (2019b) noted, an individual or interest group can become a major political player with tens of millions of dollars, suggesting that billionaires would retain ample scope to wield political influence even if they were heavily taxed. They also pointed out that the wealth tax would probably not apply to the nonprofit organizations that wealthy individuals (and others) finance to influence policy.

Most economists believe that a wealth tax could significantly reduce wealth inequality. In April 2019, the University of Chicago Booth School’s Initiative on Global Markets (IGM) Forum asked its ideologically diverse panel of 41 expert economists about their reactions to the statement, “If successfully enforced, Senator Warren’s proposed wealth tax would substantially decrease the share of wealth going to the top 0.1% of wealth-holders after 20 years.” Of the 35 economists who expressed an opinion, four strongly agreed, 19 agreed, nine were uncertain, two disagreed, and one strongly disagreed (IGM Forum, 2019).

2.3 Treatment of Wealth Under the Income Tax

Another way to collect more taxes from top wealth holders would be to increase the income taxes that they pay on the income generated by their wealth. However, increased income taxation under current income tax rules would fail to reach unrealized capital gains, which are a major type of income generated by wealth. When capital gains are realized, they are usually taxed at preferential rates, which typically also apply to dividends.

When an asset rises in value, the owner experiences economic income from the accrued capital gain, even if the gain has not been realized by selling the asset. In some cases, the owner may be able to turn the accrued gain into cash by borrowing against the appreciated asset or by using other financial strategies. Nevertheless, under the income tax system’s realization principle, income tax is generally not imposed on the capital gain until it is realized through sale of the asset. That interest-

---

3 The income tax system also fails to tax the rental value of personal residences and the value of the services provided by other consumer durable goods and collectibles.
free tax deferral reduces the owner’s tax burden because a dollar paid tomorrow is worth less than a dollar paid today. Moreover, if the owner dies without selling the asset, nobody ever pays income tax on the unrealized gain that accrued during the owner’s lifetime. Under the income tax system’s basis step-up provision, the owner’s heirs are treated as if they purchased the asset at its market value on the date of the owner’s death, so they are taxed (if they ever sell the asset) only on gains that accrued after the owner’s death.

Because many top wealth holders experience significant unrealized gains, they are taxed on only part of their economic income. Bourne et al. (2018) found that the annual income reported by top wealth holders on their income tax returns was less than 4% of their wealth. Because part of their income received the preferential rates for capital gains and dividends, their tax burden was equivalent to paying ordinary income tax rates on annual income of less than 3% of their wealth. The authors noted that the wealth holders’ total annual returns, including unrealized gains, were likely 8% or higher.4

Under current income tax rules, there is limited scope for additional taxation of the capital gains of top wealth holders. Eliminating the preferential rates on realized capital gains would still leave unrealized gains outside the tax base and might be counter-productive if it caused taxpayers to realize fewer gains (the “lock-in effect”). The wealth tax would overcome these limitations by directly taxing the asset values.

However, the income tax rules could be changed. The basis step-up rule could be replaced by a basis carry-over rule, so that when heirs (or their heirs, and so on) sell an asset, they would pay income tax on all of the gains that have accrued since the asset was originally purchased, virtually guaranteeing that gains would eventually be taxed. A more aggressive option would tax the accrued gains when the original holder died, so that the tax could never be deferred beyond the original holder’s lifetime.

More dramatic options become available if the realization principle is discarded. Under mark-to-market taxation, gains would be taxed each year as they accrued, even if the assets had not been sold. Gains could easily be taxed at ordinary income tax rates rather than the current preferential rates because the lock-in effect would no longer be an issue. Warren (2019b) proposed mark-to-market taxation of capital gains at ordinary income tax rates for the wealthiest 1% of taxpayers, in addition to a wealth tax. Because mark-to-market taxation would require asset values to be determined, it would encounter the same valuation challenges, discussed below, that wealth taxes confront. Toder and Viard (2016) proposed mark-to-market taxation of capital gains at ordinary income tax rates, though only for publicly traded assets, whose values can be easily determined. However, Grubert and Altshuler (2016) outlined an interest-charge

---

4 For corporate stock, the corporate income tax offsets the lenient individual tax treatment to some extent. However, that offset does not apply to other assets held by the wealthy.
method that would have broadly similar effects to mark-to-market taxation while largely avoiding the need to determine market values. Under the interest-charge method, the gain would not be taxed until the asset was sold or the asset holder died. At that time, however, an interest charge would be added to the tax liability based on how long the asset had been held, to approximately offset the failure to tax the gain each year as it accrued. Although asset holders would still be able to defer tax by delaying the asset’s sale, they would have to pay interest on the deferral, effectively removing its economic benefit. Market values would not need to be determined, except at death (when valuation might already be required for estate tax purposes). Grubert and Altshuler recommended using the interest-charge method for all assets.

The best approach may be to use the interest-charge method for assets that are not publicly traded while using mark-to-market taxation for publicly traded assets. In September 2019, Senator Ron Wyden (D-Oregon) proposed that approach for wealthy investors (Wyden, 2019). Like wealth taxation, a combination of mark-to-market taxation and the interest-charge method would be a fundamental change to the tax system and would be controversial. However, that approach would not require the creation of a completely new tax and it would largely avoid the need to determine the values of assets that are not publicly traded. Moreover, Viard (2014) argued that effective income tax administration may eventually require a movement away from the realization principle anyway. Batchelder and Kamin (2019) discuss mark-to-market taxation and other methods of taxing the wealthy.

Economists do not have a consensus view about whether changes to the existing tax system could substitute for wealth taxation. The IGM Forum’s April 2019 survey also asked its panel of economists about their reactions to the statement, “A public policy goal that could be accomplished with a well-enforced wealth tax could be equally accomplished with modifications to existing federal taxes—for example, revising the estate tax and/or capital gains tax.” Of the 36 economists who expressed an opinion, four strongly agreed, 14 agreed, seven were uncertain, 11 disagreed, and none strongly disagreed (IGM Forum, 2019).

2.4 Saving and Investment

Saez and Zucman make the reasonable assumption that the tax would not change the fraction of wealth that taxpayers spend each year. Under that assumption, the taxpayers’ saving, as a fraction of wealth, would fall by an amount equal to the wealth tax rate. Supporters of the wealth tax would view the taxpayers’ decline in saving, which is a slowdown in their wealth accumulation, as a feature rather than a bug. A 6% per year wealth tax would reduce the taxpayer’s wealth (relative to what it otherwise would have been) by 60% over 15 years and by 84% over 30 years. An 8% per year wealth tax would reduce wealth by 71% over 15 years and by 92% over 30 years. Saez and Zucman (2019d) tabulate the sharp reductions in top billionaires’ wealth that would have occurred if wealth taxes had been in effect since 1982.
The reduction in the taxpayers’ saving would initially equal wealth tax revenue, but the saving reduction would be larger than revenue in subsequent years. In addition to falling by the amount of the tax payment, saving would also fall because the taxpayers would have lower accumulated wealth from which to save.

As Saez and Zucman (2019b) noted, the wealth tax’s net impact on total national saving would depend on the extent to which any of the tax revenue was saved. Although some of the revenue could be saved through deficit reduction and infrastructure investment, very little of the revenue might be devoted to those purposes. Sanders (2019) stated that the revenue would be used to pay for affordable housing, universal childcare, and part of Medicare for All, while Warren (2019a; 2019b) stated that the revenue would be used for “badly needed investments in rebuilding our middle class” and to cover part of the costs of Medicare for All. Some uses of the revenue might increase human capital, a form of saving that is not included in the national income accounts.

A reduction in national saving would be financed by a reduction in investment in factories, equipment, and other capital in the United States, by a larger inflow of capital from abroad, or by a combination of both. A larger capital inflow, which represents increased borrowing from foreigners, would be manifested in a larger trade deficit. With fewer funds available from American savers to finance investment, investment must fall unless foreign savers supply more funds.

A reduction in investment in the United States would result in a smaller capital stock, making workers less productive and lowering their wages over time. Workers would then ultimately bear part of the burden of the wealth tax. Nevertheless, the decline in investment (and the wage reduction) would be ameliorated because a significant part of the saving decline would probably be financed by increased capital inflows, as Saez and Zucman (2019b) noted. The Congressional Budget Office uses a central estimate under which 57% of a decline in national saving is financed by a reduction in investment (with the other 43% financed by larger capital inflows), but also considers alternative assumptions under which the investment reduction is 71% or 39% of the saving reduction (Huntley, 2014). Only a small portion of the wealth tax burden would likely be shifted to workers.

It should be noted that alternative methods of taxing the wealthy, such as mark-to-market taxation and the interest-charge method, would also be likely to reduce national saving.

2.5 Breadth of Tax Base

Wealth taxes can have broad tax bases that cover almost all types of assets or narrow tax bases that exempt many types of assets. A broad tax base would be preferable because it would treat different assets neutrally and would raise any given amount of
revenue at a lower tax rate. Warren (2019a) called for a very broad tax base consisting of “all household assets … including residences, closely held businesses, assets held in trust, retirement assets, assets held by minor children, and personal property with a value of $50,000 or more.”

Unfortunately, the international experience suggests that it would be difficult to adopt a wealth tax with a broad base. Brumby and Keen (2018) stated, “The design of wealth taxes is notoriously prone to lobbying and the granting of exemptions that the wealthiest can exploit,” and OECD (2018) described how lobbying led to exemptions being granted under European wealth taxes. Edwards (2019) noted that many of the European taxes provided exemptions for farm assets, small businesses, pension assets, artwork, and other items. Taxpayers holding exempt assets were still allowed to deduct their full liabilities, yielding an even more distorted picture of their net worth. Leiserson, McGrew, and Kopparam (2019) provided a detailed tabulation of asset exemptions in past and present European wealth taxes. Davison (2019) predicted that various interest groups would similarly press for exemptions under a U.S. wealth tax. Saez and Zucman (2019b) countered that the Warren proposal would apply only to a small group of households with wealth above $50 million, whose pleas for asset exemptions would draw little political support. That argument seems plausible, although it may be difficult to reconcile with the contention that top wealth holders have excessive political power.

Exemptions would directly reduce the tax’s revenue yield. They would also encourage taxpayers to inefficiently shift their holdings from taxed assets to exempt assets, which would further reduce the revenue yield. Such shifting was observed in France, Germany, Norway, and Spain (Davison, 2019).

It should be noted that Congress might also add asset-specific exemptions to alternative methods of taxing the wealthy, such as mark-to-market taxation and the interest-charge method. For example, Wyden (2019) would exempt some personal residences and family farms.

2.6 Administration, Avoidance, and Evasion

Under an annual wealth tax, the fair market values of all assets and liabilities would need to be determined each year for all households with wealth (potentially) above the exemption amount. Bank accounts and publicly traded financial assets would be straightforward to value, but assets that are not publicly traded, such as land, houses, privately held businesses, artwork, and furniture, would pose difficulties. Taxpayers would have the opportunity to conservatively value, or flatly undervalue, those assets to some extent. Taxpayers might also illegally conceal assets. Moreover, taxpayers might shift their holdings toward assets that are easier to undervalue or conceal; for example, some households might move their wealth abroad because foreign assets might be easier to conceal.
Two other types of taxes, property taxes and estate and gift taxes, must also detect and value assets. However, those tax systems generally perform these tasks on a smaller scale than the wealth tax would and they often do not perform them well. Their experience therefore offers limited encouragement about the wealth tax’s ability to detect and value assets. State and local property taxes are imposed each year and apply to a vastly larger group of taxpayers than the small group that would be subjected to the wealth tax. However, property taxes primarily apply to land and structures located in the United States, thereby avoiding some of the appraisal challenges and virtually all of the concealment challenges faced by the wealth tax. Moreover, property tax appraisals are notoriously inaccurate. The estate and gift tax system must value all types of assets and it applies to a somewhat larger group of people than those subject to Warren’s proposed tax. However, the tax is imposed only when assets are conveyed by gift or bequest rather than every year; the Internal Revenue Service (IRS) processes 4,000 estate tax returns each year but would process 75,000 wealth tax returns each year under the Warren proposal and 180,000 returns each year under the Sanders proposal (Davison, 2019). Estate and gift tax valuations are also highly imperfect.

International experience has been mixed. Edwards (2019) and Davison (2019) noted that administration and compliance issues played a role in several European countries’ decisions to repeal their wealth taxes, with Rosalsky (2019) citing it as the key factor in Austria’s 1993 repeal. Saez and Zucman (2019b) reported that wealth tax avoidance and evasion were modest in Sweden and Denmark, which had extensive third-party reporting of wealth, but were more severe in Columbia and Switzerland, where enforcement was weaker.

Warren (2019a) and Sanders (2019) called for significant increases in the IRS enforcement budget, minimum audit rates for households subject to the wealth tax, and third-party reporting of financial assets based on existing international agreements. Wamhoff (2019) pointed out that a dramatic increase in IRS enforcement resources could be financed by a tiny fraction of the wealth tax revenue.

Saez and Zucman (2019e) and Wamhoff (2019) offered proposals to improve administration and compliance. For example, Saez and Zucman proposed increased information reporting on financial assets, valuing businesses based on book values of assets or by applying multipliers to annual profits, and valuing artwork by its insurance value. Wamhoff suggested that state and local governments be empowered to acquire property through eminent domain at a price equal to the value that owners reported for wealth tax purposes. Stein (2019) surveyed the outlook for wealth tax enforcement.

---

5 For decedents dying in 2020, the estate tax applies if the decedent’s estate plus cumulative lifetime taxable gifts exceed $11.58 million ($23.16 million for married couples).
Wealth Taxation: An Overview of the Issues

Most economists believe that the wealth tax would face significant administration and compliance challenges. The IGM Forum asked its panel of economists about their reactions to the statement, “Senator Warren’s proposed wealth tax would be much more difficult to enforce than existing federal taxes because of difficulties of valuation and the ways by which the wealthy can under-report their true wealth.” Of the 37 economists who expressed an opinion, nine strongly agreed, 21 agreed, four were uncertain, three disagreed, and none strongly disagreed (IGM Forum, 2019).

Individuals could legally avoid the wealth tax by emigrating and renouncing their U.S. citizenship (for simplicity, “expatriating”). Edwards (2019) noted that wealth taxes in France and Sweden prompted some high-wealth individuals to emigrate. Edwards (2019) and Davison (2019) reported that emigration helped prompt France to repeal its tax. However, expatriation by Americans is generally more difficult than moving within the European Union. Warren (2019a) also called for an “exit tax” equal to 40% of wealth in excess of $50 million on Americans who expatriate and Sanders (2019) proposed a similar tax, but with a 60% tax rate on wealth in excess of $1 billion. Such an exit tax would likely deter many potential expatriations and would offset part of the revenue loss that would arise from any remaining expatriations. As Stein (2019) noted, the exit tax could be applied retroactively to individuals who expatriated while the wealth tax was being considered but before it was enacted.

Taxpayers could also avoid the tax by giving assets to relatives (other than spouses and unmarried minor children) whose wealth was below the exemption amount. Although such gifts would reduce the tax’s revenue yield, Saez and Zucman (2019e) argued that they would advance the tax’s goal of breaking up concentrated wealth. They also acknowledged that the tax could be avoided by giving wealth to foundations and other charities, but argued that such gifts would be socially beneficial.

2.7 Revenue Yield

Saez and Zucman (2019a) estimated a $2.75 trillion revenue yield over 10 years from the original Warren (2019a) proposal, including $212 billion in the first year. Warren (2019b) stated that the modified proposal would raise an additional $1 trillion over 10 years. Saez and Zucman (2019d) estimated a $4.35 trillion revenue yield over 10 years from the Sanders proposal, including $335 billion in the first year. Their estimates allowed for a 15% revenue loss from tax avoidance and evasion. It seems likely, however, that avoidance and evasion would be larger under the Sanders (2019) and Warren (2019b) proposals, due to their higher tax rates. Smith, Zidar, and Zwick (2019) estimated lower revenue yields for the original Warren proposal based on their lower estimates of wealth concentration. Their estimates also found that a larger portion of wealthy taxpayers’ holdings are in assets that are more difficult to value, suggesting greater scope for tax avoidance.

Using a different methodology, Summers and Sarin (2019a) obtained a radically lower revenue estimate for the original Warren proposal, highlighting the important
role played by the breadth of the tax base and the scope for evasion and avoidance. Based on mortality data, they estimated that the once-in-a-lifetime, 40% estate tax is equivalent to a 0.8% per year wealth tax. Because the estate tax raises $10 billion per year from estates larger than $50 million, they estimated that a 2% per year annual tax on wealth above $50 million would raise $25 billion per year, approximately one-eighth of the Saez-Zucman estimate. Although they acknowledged that some upward adjustments to their estimate might be warranted, they concluded that “it is likely extremely premature to bank on anything like the $200 billion plus that Saez and Zucman estimate.” Saez and Zucman (2019c) forcefully argued that the $25 billion Summers-Sarin number was far too low, noting that the tax would raise more than that from the 15 wealthiest Americans alone. They argued that most assets held by the very wealthy are relatively easy to value and emphasized the enforcement provisions of the Warren proposal. Summers and Sarin (2019b) acknowledged again that upward adjustments to their $25 billion number were warranted, but insisted that the experiences with the estate and gift tax and European wealth taxes suggested that revenue would fall short of the Saez-Zucman estimate. Saez and Zucman (2019e) argued that the Summers-Sarin number was also biased downward by inaccurate assumptions about mortality rates and by a failure to account for estate tax deductions that would not be available under the wealth tax.

By basing their estimate on the current estate tax, Summers and Sarin effectively assumed that Congress would add to the wealth tax the same type of base-narrowing provisions that it has adopted under the estate tax and that taxpayers would be able to use the same types of strategies to avoid the wealth tax that they use to avoid the estate tax. They argued that those assumptions were likely to hold. They also noted that many tax proposals end up raising much less revenue than a simple analysis of macroeconomic data would suggest. However, Gene Sperling, former economic adviser to Presidents Bill Clinton and Barack Obama, countered that the “miserable state of enforcement of the estate tax” could be “improved with smart public policy” and should not be treated “as an immovable part of nature” (Schor, 2019).

According to the Saez-Zucman estimates, wealth tax revenue would be approximately 1% of GDP under the original Warren proposal and 1.6% of GDP under the Sanders proposal. Those revenue yields would be high relative to European wealth taxes. Leiserson, McGrew, and Kopparam (2019) reported that Norway’s tax raises 0.4% of GDP and Saez and Zucman (2019b) noted that Spain’s tax and France’s former tax raised 0.2% of GDP. However, Switzerland’s tax raised approximately 1% of GDP. Davison (2019), Edwards (2019), and Rosalsky (2019) reported that disappointing revenue yields played a role in some European countries’ decisions to repeal their wealth taxes.

Due to the wealth tax’s novelty, its revenue yield is difficult to determine. On balance, it is probably reasonable to assume that revenue would fall somewhat short of the Saez-Zucman estimates. Despite Senator Warren’s commendable embrace of a broad tax base, Congress would likely narrow the tax base to some extent, in accord with European practices and its own estate tax practices.
3. Constitutional Questions

A federal wealth tax would face potential constitutional challenges. The original Constitution required that all “direct” federal taxes be apportioned among states in proportion to their population, although the Sixteenth Amendment, adopted in 1913, exempted income taxes from that requirement. If the wealth tax were apportioned, the tax rate would be lower in states with higher per-capita wealth in order to equalize per-capita tax liabilities across states. That rate differentiation would be a severe flaw, making an apportioned wealth tax unattractive.

The wealth tax would escape the apportionment requirement if it was either an indirect tax or an income tax. The classification of the tax would depend upon unresolved legal issues and the tax’s features.

It is generally understood that a tax on real property would be a direct tax and would have to be apportioned. The U.S. Supreme Court ruled in 1796 that a tax on personal property (in that case, carriages) was an indirect tax that need not be apportioned (Hylton v. United States). The Court ruled in 1881 that income taxes were indirect taxes that need not be apportioned even though income from real property was in the tax base (Springer v. United States). However, the Court overruled that decision in 1895 (and backed away from its 1796 decision), holding that taxes on income from either real or personal property were direct taxes that had to be apportioned (Pollock v. Farmers’ Loan & Trust Co.). In later decisions, though, the Court moved toward a narrower definition of direct taxes. It ruled in 1900 that the estate and gift tax was an indirect tax imposed on the privilege of conveying property by gift or bequest rather than a direct tax on property and therefore did not need to be apportioned (Knowlton v. Moore). Similarly, it ruled in 1910 that the corporate income tax was an indirect tax imposed on the privilege of operating in corporate form rather than a direct tax on income from property and therefore did not need to be apportioned (Flint v. Stone Tracy). Of course, the adoption of the Sixteenth Amendment in 1913 made it irrelevant whether income taxes are direct. In 2012, the Supreme Court ruled that an annual tax on certain persons not covered by health insurance was an indirect tax that need not be apportioned (National Federation of Independent Business v. Sebelius).

It is difficult to discern from the Court’s decisions whether a wealth tax would be a direct tax. The tax base includes real and personal property, with a deduction for liabilities. When she introduced her proposal, Senator Warren released two letters, Ackerman et al. (2019) and Johnsen et al. (2019), from 17 law professors stating that the tax would be an indirect tax that need not be apportioned. Johnsen and Delinger (2018) provided a more complete exposition of that position and Wamhoff (2019), Feldman (2019), and Thornton and Hendricks (2019) also argued that a wealth tax would probably be an indirect tax. Other commentators were less sanguine. Jensen (2003), Freeman (2019), and Khan (2019) argued that the wealth tax would be a direct tax that would need to be apportioned. Bishop-Henchman (2019) noted that
the issue was unclear, but said that he was inclined to think that the wealth tax would be a direct tax. Barro (2019) surveyed the uncertainty, concluding that the wealth tax would face a significant risk in court. Sarin and Summers (2019b) argued that it would be dangerous to put significant political effort into a tax that the courts might strike down as unconstitutional.

Even if a straight wealth tax would be a direct tax that would have to be apportioned, suitable modifications might transform it into either an indirect tax or an income tax. Glogower (2019) proposed that high-wealth households be required to make tax payments that would be labeled as additional income taxes rather than as wealth taxes, despite being based on wealth; it is unclear whether the courts would accept that disguise. It also might be possible to label the wealth tax as an income tax on presumed income from wealth, as the Netherlands does, or to treat the wealth tax as an advance payment of estate and gift taxes.

A final option would be to reluctantly accept apportionment. Buchanan (2019), arguing that an apportioned wealth tax would be better than none, proposed that the wealth tax legislation include a fallback provision that would institute an apportioned tax if the courts ruled that the unapportioned tax was unconstitutional. He also conjectured that the courts might be more reluctant to strike down the unapportioned wealth tax if they knew that it would automatically be replaced by an apportioned wealth tax, a replacement that nobody would welcome.

It should be noted that mark-to-market taxation could also face constitutional challenges. However, mark-to-market taxation would run less risk of being struck down than a wealth tax. The challenges to mark-to-market taxation would be based on the argument that the Sixteenth Amendment’s exemption of income taxes from the apportionment requirement applies only to taxes on realized income. Miller (2014) notes that most legal scholars reject this argument and concludes that the courts are unlikely to embrace it.

4. Conclusion

Annual wealth taxation is one strategy for taxing extremely wealthy households, including those who defer or escape income tax on their unrealized capital gains. However, a wealth tax would pose administrative and constitutional challenges. Although those challenges could probably be overcome if necessary, it may be prudent to pursue any increased taxation of the affluent through other policies that would not pose the same difficulties. A number of commentators who favor increased taxation of the rich, including Sarin and Summers (2019a), the Washington Post (2019), and Hemel (2019), persuasively argued that it would be better to pursue reform of the income tax and estate and gift taxes.
Sarin and Summers outlined a package of progressive changes to the income tax system that they estimated would bring in $2.83 trillion over 10 years, slightly more than the Saez-Zucman estimate of the original Warren proposal’s revenue yield. Revenue estimates for the Sarin–Summers proposals are likely to be more reliable because they are reforms of the existing tax system rather than a completely new tax. The proposals’ actual revenue yield would therefore likely be close to their estimated yield, while, as discussed above, the wealth tax’s actual revenue yield might be significantly lower than the Saez-Zucman estimate.

One of the Sarin–Summers proposals would replace basis step-up with basis carry-over, so that, as explained above, capital gains not realized during a taxpayer’s lifetime would be taxed when an heir eventually sold the appreciated asset. However, it might be desirable to go further, adopting mark-to-market taxation for capital gains on publicly traded assets and a deferral charge with taxation at death for assets that are not publicly traded, while taxing capital gains at ordinary income tax rates.

Although the changes proposed by Sarin and Summers, and other possible changes within the income tax system, would primarily be borne by affluent taxpayers, they would not fall exclusively on the very wealthiest households. Accordingly, the proposals would not break up the concentration of wealth to the same extent as the wealth taxes proposed by Senators Warren and Sanders. Some wealth tax supporters might therefore find them a disappointing substitute. As discussed above, however, there is little reason to think that breaking up the concentration of wealth would have much impact on the distribution of political power in the United States. And significant increases in progressivity can clearly be achieved without imposing the entire burden on a tiny group of extremely wealthy households.

Although the wealth tax is a bold proposal, bold is not always better.
References


*Flint v. Stone Tracy*, 220 U.S. 107 (1910)


*Hylton v. United States*, 3 U.S. 171 (1796)


*Knowlton v. Moore*, 178 U.S. 41 (1900)


*Springer v. United States*, 102 U.S. 586 (1881)


Policy Options for Taxing the Rich

AUTHORS
Lily Batchelder* and David Kamin†

ABSTRACT
The U.S. economy exhibits high inequality and low economic mobility across generations relative to other high-income countries. The United States will need to raise more revenues in order to reduce these disparities, finance much-needed new services and investments, and address the nation’s long-term fiscal needs. This paper outlines policy options for raising a large amount of revenues primarily from the most affluent, first discussing potential incremental reforms and then focusing on four main options for more structural reform: dramatically increasing the top tax rates on labor and other ordinary income; effectively taxing the wealthy on accrued gains as they arise and at ordinary rates; a wealth tax on high-net-worth individuals; and a financial transactions tax. Although we summarize the relative advantages and disadvantages of these approaches, we generally conclude that they all merit serious consideration. Several options are also complementary to one another. In practice, however, the relative strengths of each of these policies will depend to a large extent on how each is designed after it has made its way through the legislative and regulatory process.

* Robert C. Kopple Family Professor of Taxation, New York University School of Law; lily.batchelder@nyu.edu.
† Professor of Law, New York University School of Law; kamin@nyu.edu.
For helpful comments and discussions, we are grateful to Kim Clausing, Seth Hanlon, Janet Holtzblatt, Chye-Ching Huang, Samantha Jacoby, Bill Gale, Amy Ganz, Itai Grinberg, Greg Leiserson, Eric Toder, Kees van Raad, Alan Viard, Jim Wetzler, and Eric Zwick. We thank Jay Cullen, Scott Greenberg, and Yuyao Leng for excellent research assistance. All errors are ours.
1. Introduction: Why Raise More Revenues From the Wealthy?

The United States has one of the highest levels of income and wealth inequality among high-income nations. As illustrated in Figure 1, the United States has the second highest level of income inequality after taxes and transfers among 33 Organisation for Economic Co-operation and Development (OECD) countries. It has the highest level of wealth inequality in the OECD (Balestra & Tonkin, 2018). Income and wealth inequality are heavily skewed by race. Average net worth for blacks in the United States is only 14% of that for non-Hispanic whites (Wolff, 2018). As a group, the top 1% in the United States receives more income than the bottom 40% and owns more wealth than the bottom 95% (Wolff, 2017).

Further, the United States has one of the lowest levels of intergenerational economic mobility among high-income countries. On average, a father in the United States passes on roughly half of his economic advantage or disadvantage to his son (Corak, 2013). Among other high-income countries, the comparable figure is typically about one-third, and in several countries it is one-fifth. There are even larger mobility barriers among some communities of color. Black men in particular have far less upward mobility and greater downward mobility than others, and to such a large extent that the current black-white income gap is not projected to change at all if these mobility dynamics persist (Chetty, Hendren, Jones, & Porter, 2018).

Thus, to an unusually large extent in the United States, economic disparities between individuals reflect the luck of one’s birth and systemic discrimination, not hard work.

**Figure 1. Income Inequality After Taxes and Transfers in OECD Countries**

![Figure 1. Income Inequality After Taxes and Transfers in OECD Countries](source)

**Source:** Authors’ calculations based on OECD (2019). We use the most recent year available from 2014 to 2016 and exclude countries for which no data on transfers is available. See Huang and Frentz (2014) for discussion of the OECD methodology. The Gini coefficient is a measure of income inequality.
One of the ways we can begin to address these vast disparities of income and opportunity is through a more progressive fiscal system that expands much-needed investments in individuals who do not benefit from such privileges.

In addition, the United States faces a long-term fiscal shortfall. There is reason to believe that debt may pose fewer risks to the economy than it has in the past, given the long-term decline in interest rates and the substantial fiscal capacity in the United States and other high-income economies (Blanchard, 2019; Furman & Summers, 2019). Nonetheless, most economists agree that markets would eventually react if debt grew continuously as a share of the economy, which current projections suggest it will absent policy changes (Gale, 2019).

For instance, the Congressional Budget Office (CBO) projects that, under current law, the federal debt will rise from about 80% of Gross Domestic Product (GDP) today to almost 150% by 2050 (CBO, 2019). Stabilizing the debt-to-GDP ratio over this period will require reducing annual deficits by about 2% of GDP—and more than double this if expiring provisions, which include much of the 2017 tax cuts, are continued.¹

Despite these pressing needs, U.S. federal revenues as a share of GDP were 16.5% in 2018. This is 1 percentage point below the average over the past 50 years, even though we are in the midst of an economic expansion (CBO, 2019). As of 2017, total U.S. revenue as a share of GDP was already 7 percentage points below the OECD average, and that was before the 2017 bill’s tax cuts took effect (OECD, 2019c).

Theoretically, it is possible to raise sufficient revenues to stabilize our fiscal outlook and create a more progressive fiscal system with tax increases that apply across the economic distribution, not just to the wealthy. For example, every other high-income country in the world has a federal value-added tax (VAT) (Tax Policy Center, 2019a). If the United States followed the model of other high-income countries of raising much more revenue from broad-based taxes, principally a VAT, and spending it in relatively progressive ways, economic disparities would decline, though not necessarily between the upper-middle class and the wealthy.

But if history is any guide, the United States may not follow this model. We redistribute relatively little through our fiscal system, ranking 28 out of 33 OECD countries.² But to the extent we do, we are the only OECD country that relies roughly equally on the tax system and direct spending programs to mitigate economic disparities (Joumard, Pisu, & Bloch, 2012). All others rely more heavily on direct spending programs, and often dramatically so.

If the United States persists with its relatively tax-focused model of redistribution, reducing economic disparities and stabilizing our fiscal outlook will requiring raising

¹ Authors’ calculations based on CBO (2019).
² Author’s calculations based on OECD (2019).
substantial new revenue from measures that focus to a large extent on the affluent, rather than the public at large. Accordingly, this essay focuses on the pros and cons of different options for raising substantial additional revenue over the next decade primarily from the wealthy.

2. What Is Wrong With the Current System?

Beyond raising insufficient revenues, there are a number of serious problems with the current U.S. approach to taxing wealthy individuals. Broadly speaking, the current system offers too many ways for those with the greatest resources to escape tax, either by reducing their effective tax rate or avoiding taxes altogether. The result is that, while some at the top are taxed at the highest rates, many are not. Absent changes to the tax base that make it harder to avoid taxes, many wealthy individuals would be largely unaffected by increases to the top statutory tax rates.

2.1 How Those at the Top Differ

Those at the top earn income and accrue wealth in fundamentally different ways than the rest of the population. Wages comprise the vast majority of income for those outside of the top 1% of income. Tax avoidance and evasion are rare for wage income because it is subject to information reporting and withholding, and because wage earners generally cannot manipulate the timing of income recognition (Slemrod, 2007).

But those at the top earn or report their income differently. Much more of their income comes from capital gains, dividends, and income flowing through business entities. These forms of income are often eligible for preferential rates. For example, in 2016, wages and salaries comprised only 10% of the income of the top 0.001%, while capital gains and dividends taxed at preferential rates made up 71% of their income, with business income comprising the remainder (Batchelder & Kamin, 2019).

Further, the actual share of income that those at the top derive from capital gains and business income may be substantially higher. That is because it is easier to legally defer paying tax on such income, or even eliminate the tax on it entirely. For instance, as explained in more detail below, gains on property only have to be reported when they are “realized”; that is, when property is sold or exchanged. Gains on property are also more often (illegally) underreported because they are not subject to withholding or, sometimes, information reporting. Thus, the data we use here, which rely on reported income from tax returns, understates the share of income that the wealthy earn from these sources.
2.2 Tax Avoidance Strategies Used by the Wealthy

High-income Americans and the entities they own make use of a number of tax avoidance strategies, which we describe in more detail in Batchelder and Kamin (2019). Many involve shifting income from a high tax rate category to a low or zero tax rate category. Of course, there are often costs of making such shifts, such as lawyer fees and difficulties in restructuring transactions, so not all income ends up in the lowest rate category. But, high-income Americans still have plenty of opportunities to engage in such shifting.

There are several reasons why the wealthy are able to pay tax at much lower rates than the headline tax rates imply, and the 2017 tax law made this problem worse.

First, the wealthy tend to characterize a large share of their income from labor as income from business entities, which now can be subject to much lower rates even if it is ordinary income. Smith, Yagan, Zidar, and Zwick (2016) estimate that three-quarters of the business profits received by the wealthy derive from their human capital, not physical or financial capital.

Before the 2017 law, characterizing labor income in this way—as ordinary income being earned through a business—could achieve some real but relatively limited tax savings for the wealthy. The 2017 law substantially expanded opportunities to use business entities to avoid the top rate (for further discussion, see Kamin et al., 2019).

The wealthy now have two main options for using business entities to achieve substantial tax savings relative to being taxed at the top rate. One is to characterize their income as earned through a pass-through business entity—the vast majority of which is eligible for a 20% deduction under the 2017 tax law (JCT, 2019; Goodman, Lim, Sacerdote, & Whitten, 2019). This can bring their top marginal rate down from 40.8% to 29.6% if they also take advantage of loopholes in the Medicare self-employment tax (SECA) and the net investment income tax (NIIT). Alternatively, they may claim that their income—including their labor income—is earned by a business they own that is subject to the corporate income tax (so-called C-corporations). In the wake of the 2017 tax law, the top tax rate on such income is only 21%, down from 35%. Wealthy individuals who report income through a C-corporation do have to pay personal income taxes on dividends or realized capital gains on their stock at a top rate of 23.8% (including the NIIT). But this tax can be deferred indefinitely or eventually disappear if the stock is held until death or another loophole is used to escape the second layer of tax.

Second, the wealthy who elect the pass-through route can often claim their income, including their labor income, as long-term capital gains or dividends. In this case, the top rate falls from 40.8% to 23.8%. If they take advantage of SECA and NIIT loopholes, they can further lower their top tax rate to 20%. Carried interest—available to managing partners in private equity and similar industries—is just a small example of this widespread practice of converting income from labor into lower-taxed capital gains.
Third, the wealthy can often afford to defer realizing capital gains. If they do so for a long enough period of time, the present value of their top tax rate on such gains approaches zero. And if they do so until their death, the top tax rate is actually zero, thanks to a provision called “stepped-up basis,” which forgives tax on such accrued gains. Their top tax rate can also be zero if they donate appreciated property to a charity like a family foundation, even if they maintain some degree of control over it. Further, to the extent the wealthy do realize gains on some property, they can choose to sell other property with built-in losses to offset those gains.

Fourth, multinational corporations—whose owners are disproportionately wealthy—can achieve very low tax rates by exploiting differences in tax rates across international boundaries. These corporations report large amounts of income in tax havens (Zucman, 2014; Clausing, 2019a).

Fifth, while a large share of the income of the wealthy is derived from labor income, a substantial share is also the product of inheritances. Inherited income is entirely excluded from both the income tax and payroll tax bases. The estate tax and related wealth transfer taxes were meant to partially address this omission. But the exemptions are so large ($22.8 million per couple in 2019) and the base so porous that income in the form of inheritances was taxed at an average rate of less than 4% in 2009, and is taxed at even lower rates today (Batchelder, 2009).

Finally, enforcement of the existing tax laws governing the wealthy is weak and getting weaker. The audit rate for the top 1% has declined dramatically—by about 80% since 2011—and is only 1.6% today (Kiel, 2019).

All of this means that the wealthy are taxed at a wide range of rates, depending on how they report their income. There is considerable evidence that high-income Americans, and the entities they own, take advantage of this menu of tax planning options to substantially reduce their tax bill. The exact magnitude is hard to quantify because much avoidance is simply unmeasured, at least directly, or is not considered a tax underpayment because it is legal under current law. Nevertheless, an array of evidence points to very large magnitudes of foregone revenue. We summarize some of this evidence in Batchelder and Kamin (2019).

3. Selected Revenue Options Within the Current System

Within the basic structure of the current tax system, policymakers have proposed a range of policies that would raise considerable revenue from those with the greatest resources. In Table 1, we list several of these proposals to provide a sense of scale. This section is not a comprehensive compilation of all such measures, as there are many.

All of the proposals listed focus either solely or disproportionately on those with the greatest resources or the businesses they own. For organizational purposes, the table is broken down between direct repeal or reform of elements of the 2017 tax legislation,
along with further measures that could be taken. A number of these proposals would, in addition to raising revenue in a progressive fashion, reduce complexity and wasteful tax planning. We consider many to be good ideas. But since they have, for the most part, been discussed in other contexts and do not involve fundamental shifts in the system, we do not delve into the details or relative pros or cons of each here.

**Table 1: Incremental Revenue Measures**

<table>
<thead>
<tr>
<th><strong>2021-2030 (BILLIONS)</strong></th>
<th>CURRENT LAW</th>
<th>CURRENT POLICY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Repeal or Reforms of 2017 Tax Law</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Return Top Individual Rate to 39.6% from 37% (1)</td>
<td>$90</td>
<td>$200</td>
</tr>
<tr>
<td>Reverse Doubling of Estate Tax Exemption (back to $11.4M per couple) (2)</td>
<td>$60</td>
<td>$110</td>
</tr>
<tr>
<td>Repeal Pass-Through Deduction (2)</td>
<td>$280</td>
<td>$620</td>
</tr>
<tr>
<td>Increase Corporate Rate to 28% from 21% (2)</td>
<td>$730</td>
<td></td>
</tr>
<tr>
<td>Raise Minimum Tax on Foreign Income to 21% + Apply Per Country (3)</td>
<td></td>
<td>$340</td>
</tr>
<tr>
<td><strong>Sub-Total</strong></td>
<td>$1,500</td>
<td>$2,000</td>
</tr>
<tr>
<td><strong>Additional Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10% Surtax on AGI Above $2 Million (4)</td>
<td>$610</td>
<td></td>
</tr>
<tr>
<td>Tax Accrued Gains at Death and Increase CG/Dividends Rate to 28% (5)</td>
<td>$290</td>
<td></td>
</tr>
<tr>
<td>Broaden Base of Self-Employment Tax + 3.8% ACA Surtax (5)</td>
<td>$280</td>
<td></td>
</tr>
<tr>
<td>Cap Value of Itemized Deductions at 28% (6)</td>
<td>$410</td>
<td>$310</td>
</tr>
<tr>
<td>Estate Tax: $7M Per Couple Exemption, 45%-65% Rate, Limit Avoidance</td>
<td>$310</td>
<td></td>
</tr>
<tr>
<td>Return to 2009 Parameters + Anti-Avoidance Measures (5)</td>
<td>$210</td>
<td></td>
</tr>
<tr>
<td>Increase Rates on Largest Estates (Max = 65% on Transfers &gt;$1B) (7)</td>
<td>$100</td>
<td></td>
</tr>
<tr>
<td>Eliminate Accelerated Cost Recovery for Largest Businesses (2&amp;8)</td>
<td>$760</td>
<td>$920</td>
</tr>
<tr>
<td><strong>Sub-Total</strong></td>
<td>$2,970</td>
<td>$3,030</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>$4,470</td>
<td>$4,970</td>
</tr>
<tr>
<td><strong>% of GDP</strong></td>
<td>1.6%</td>
<td>1.8%</td>
</tr>
</tbody>
</table>

**Sources are authors’ calculations based on:** (1) AEI Tax Brain; (2) JCT (2017, 2018); (3) Clausing (2019b); (4) Tax Policy Center (2019c), (5) JCT (2016); (6) JCT (2011) and Tax Policy Center (2018); (7) Auxier, Burman, Nunns, & Rohaly (2016) and Sammartino, Burman, Nunns, Rosenberg, & Rohaly (2016); Batchelder (2017a). The authors have updated all estimates to be consistent with a 2021-30 budget window, with details provided in Batchelder and Kamin (2019).
In presenting these and other revenue estimates, we use the 10-year budget window for the next Congress (2021–2030) and assume each proposal is effective immediately. (See Batchelder and Kamin, 2019, for details as to how these estimates are derived.) Where relevant, we present the revenue raised relative to both “current law” and “current policy.” Under current law, many of the 2017 tax law’s provisions affecting individual income taxes expire after 2025. Thus, several measures (such as increasing the top rate back to the previous top rate) raise revenue only temporarily relative to current law. By contrast, if measured relative to an alternative scenario in which the tax cuts are continued, these measures raise more. Other measures, such as limiting itemized deductions, raise more relative to current law than current policy since the tax law already contains limitations on these deductions that are scheduled to expire.

Some who support raising taxes on the wealthy think we should maintain the basic structure of the current system but reform it, such as in the ways listed above (e.g., Sarin & Summers, 2019a). As the table shows, these measures could raise $4.5 to $5 trillion over the decade, or 1.6% to 1.8% of GDP. This is, of course, a considerable amount of revenue.

However, it may be insufficient to address long-term, fiscal shortfalls if we maintain our existing spending commitments and, even more so, if we address significant needs for additional services and investments. As noted above, CBO projections suggest that, under current law, annual deficits will need to be reduced by almost 2% of GDP to stabilize the debt-to-GDP ratio over the next 3 decades. More than double this amount will be necessary if a number of current policies, such as the 2017 tax cuts and relief from the sequester, are continued (CBO, 2019). Significant additional revenues can be raised from those at the top, but it will tend to require the kinds of structural changes discussed in the next section.

All of these proposals would fall primarily on the wealthy but not all of them would exclusively burden the wealthy. For example, Treasury and the Joint Committee on Taxation (JCT) estimate that 75% to 82% of the burden of the corporate tax falls on corporate equity owners or owners of all capital, while 18% to 25% falls on labor (Cronin, Lin, Power, & Cooper, 2012; JCT, 2013; for further discussion of the incidence of the corporate tax, see Clausing, 2012; Batchelder, 2017b). More recent estimates by Treasury imply the burden on labor is only 12.5% in general and 7.5% in the case of multinational corporations (Power & Frerick, 2016). While capital and corporate equity ownership are highly concentrated among the wealthy, the bottom 99% still receive roughly half of all capital income (Cronin et al., 2012). Pass-through income is even more concentrated, and Treasury and JCT estimate that an even smaller portion of taxes on pass-through businesses fall on labor (Cronin et al., 2012; JCT, 2013). Nevertheless, the burden on labor is not zero. Thus, the proposals that would raise revenue through business taxes, which total $2.1 to $2.6 trillion, would fall very disproportionately on the wealthy, but a portion would be borne by middle-income investors and, to a much smaller degree, workers.
Further, while these reforms would address some of the problems summarized in the prior section, other problems would remain, and might even be exacerbated. Capital gains and dividends still would be taxed at much lower rates than income from labor and the differential would widen, increasing the pressure on the line between the two. Repeal of stepped-up basis would eliminate one major incentive to defer realizing gains. But large incentives to defer realizing gains would remain, including those due to the time value of money, potential future rate decreases, and the tax exemption for gains on property donated to charity.

A more robust estate tax would better address the direct effects of inherited advantage. But it would have smaller effects on many of the indirect advantages associated with wealth, such as social connections with other wealthy individuals, access to the best educational opportunities, and the like.

In addition, while all of these proposals maintain the basic structure of the current tax system to some degree, these changes are not necessarily more politically feasible than the structural reforms that are described in the following section. Whether “incremental” or “structural,” there will always be strong and organized opposition to such measures, and some structural changes arguably could garner stronger public support than more incremental reforms. Thus, we distinguish between incremental and structural reforms as a way of describing the degree of substantive change in the structural underpinnings of the tax system, and not of the ease or probability of enacting such reforms.

Finally, the list in Table 1 is not definitive or comprehensive, but it is intended to contain most of the incremental steps that we know of that are estimated to raise substantial revenue. While one could surely offer some other combination of such measures, the overall revenue is likely to be in the same general range as these—roughly 1% to 2% of GDP in additional revenue.
4. Options for Structural Changes to Raise Revenues From the Wealthy

This section discusses four potential structural changes to the tax system that would raise revenue primarily from those at the top: dramatically increasing the top tax rates on labor and other ordinary income; taxing accrued gains as they arise and at ordinary rates; implementing a wealth tax; and enacting a financial transactions tax. These reforms are not mutually exclusive, and several are complementary to one another. Nonetheless, we discuss their relative advantages and disadvantages.

4.1 Dramatically Raising Top Rates on Labor and Other Ordinary Income

Over two-thirds of the reported income of the top 1% is taxed at ordinary rates. As a result, dramatically increasing the top ordinary rate can generate substantial revenues. Increasing the top individual rate to 70% on income over $10 million (the top 0.01% of households), as Representative Alexandra Ocasio-Cortez has suggested, would raise about $260 to $320 billion over 10 years.

If the threshold were lower, such a dramatic rate increase would raise far more. To give a sense of scale, the ordinary income tax base above $1 million (the top 0.2% of households) is about six times larger than it is above $10 million. The ordinary income tax base for the current top bracket (income above $612,000 if married or $510,000 if single; the top 0.6% of households) is about nine times larger.

An alternative is to raise the top income tax rate somewhat less, and subject earnings above $250,000 to the Social Security tax. This latter proposal would raise roughly $1.4 trillion over a decade if enacted on its own, as summarized in Table 2. It would raise less if combined with a top income tax rate increase due to interaction effects.

---

3 Authors’ calculations from Statistics of Income, Table 3, for Tax Year 2016.

4 We acknowledge that this is less of a structural reform than the others described in this section. Nevertheless, we found it helpful to discuss it here in order to compare it to the other options.

5 We have adjusted the Penn Wharton Budget Model estimates to cover the 2021–2030 period rather than an earlier budget window. This is the case with all estimates cited in this paper.
### Table 2. Structural Reforms

<table>
<thead>
<tr>
<th></th>
<th>2021-2030 (BILLIONS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CURRENT LAW</td>
</tr>
<tr>
<td><strong>Significantly Raise Top Rates on Labor and Ordinary Income</strong></td>
<td></td>
</tr>
<tr>
<td>Increase top individual rate to 70% from 37% for income over $10M (1)</td>
<td>$260</td>
</tr>
<tr>
<td>Eliminate maximum earnings threshold in Social Security tax above $250K in earnings (2)</td>
<td></td>
</tr>
<tr>
<td><strong>Financial Transactions Tax</strong></td>
<td></td>
</tr>
<tr>
<td>0.1% tax on all financial assets (2)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2021-2030 (BILLIONS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TAX AVOIDANCE RATE</td>
</tr>
<tr>
<td></td>
<td>0%</td>
</tr>
<tr>
<td><strong>Accrual Tax</strong></td>
<td></td>
</tr>
<tr>
<td>Limited to Top 1%</td>
<td></td>
</tr>
<tr>
<td>Mark-to-market for publicly traded assets (3)</td>
<td>$2,200</td>
</tr>
<tr>
<td>Retrospective accrual tax for illiquid assets (3)</td>
<td>$600</td>
</tr>
<tr>
<td>Total</td>
<td>$2,800</td>
</tr>
<tr>
<td>Limited to Top 0.1%</td>
<td></td>
</tr>
<tr>
<td>Mark-to-market for publicly traded assets (3)</td>
<td>$800</td>
</tr>
<tr>
<td>Retrospective accrual tax for illiquid assets (3)</td>
<td>$200</td>
</tr>
<tr>
<td>Total</td>
<td>$1,000</td>
</tr>
<tr>
<td><strong>Wealth Tax</strong></td>
<td></td>
</tr>
<tr>
<td>2% tax on wealth for top 0.1% and 3% on wealth over $1B (3)</td>
<td>$3,300</td>
</tr>
<tr>
<td>2% tax on wealth for top 1% (3)</td>
<td>$6,700</td>
</tr>
</tbody>
</table>

Sources are authors’ calculations based on: (1) Ricco and Prisinzano (2019) (averaging their three estimates accounting for avoidance); (2) Congressional Budget Office (2018); (3) Survey of Consumer Finance and other sources. For more details, see Batchelder and Kamin (2019).

#### 4.1.1 Advantages

This is of course a large amount of revenue, which would be raised almost exclusively from those who are well-off. While no tax solely burdens the individuals remitting it, taxes that are directly limited to affluent individual taxpayers tend to be shifted on to others to a lesser extent than taxes that only indirectly focus on the affluent, such as corporate income taxes. Eliminating the maximum earnings threshold in the Social Security tax would also help to stabilize the Social Security trust fund by delaying its exhaustion for an additional 13 years (CBO, 2018).
There is precedent for such high individual income tax rates in the United States and abroad. Indeed, from 1936–1981, the top ordinary rate in the United States was 70% or higher (Tax Policy Center, 2019b). This approach also would not entail the valuation and liquidity challenges associated with some of the other potential structural reforms. Finally, there is at least some evidence that at high marginal tax rates, those with the highest incomes engage in less “rent-seeking” behavior, which could both reduce such wasteful activity and redistribute income down the income spectrum (Piketty, Saez, & Stantcheva, 2014). However, there are several potential downsides.

### 4.1.2 Potential Challenges

At high tax rates, there are greater incentives for earners to change their behavior to reduce taxes, whether through changes in real economic transactions or how income is reported. The degree of these responses depends on the underlying legal rules and the tax rates that are applied to other tax bases. Thus, broadening the tax base and harmonizing tax rates on other forms of income should be seen as an important complement to significant marginal rate increases on any given type of income.

On their own, these reforms would dramatically increase the already large difference between the tax rates on labor or ordinary income and those on capital income, including capital gains and dividend income. As a result, raising the top tax rate would substantially increase incentives for the wealthy to recharacterize labor and ordinary income as one of the other, lower-taxed categories of income. Further increasing opportunities for tax avoidance would, in turn, render the tax system less efficient, more complex, and, at least among the wealthy, less fair.

But these real downsides could be largely addressed if the taxation of capital were reformed to apply similar rates to capital gains and dividends in a manner that raised revenue, such as through the accrual-based tax system described next.

### 4.2 Accrual Tax

Unfortunately, if no other rules are changed, raising the rate on capital gains and dividends to the same level as ordinary income would likely lose revenues relative to some lower rates on such income, at least as estimated by the JCT and Treasury. They assume—again if no other rules are changed—that the capital gains rate that maximizes revenues is in the range of 30% because of the lock-in effect. That is, above a tax rate on capital gains of roughly 30%, the Treasury would begin to lose revenues because taxpayers would respond by deferring realizing gains for much longer periods of time. Some believe the revenue-maximizing rate on capital gains is higher than the JCT and Treasury assume (Gravelle, 1991). But there are

---

6 Authors’ calculations based on Congressional Research Service (2019).
also relatively few countries that tax long-term capital gains at rates above 30% and whose experience could therefore be used to empirically examine what the revenue-maximizing rate actually is. Notable exceptions include Austria, Chile, Denmark, Finland, Ireland, South Korea, Thailand, and Turkey, some of which tax capital gains at much higher rates than 30% (Ernst & Young, 2018).

If taxing capital gains and ordinary income at the same rates were coupled with certain other reforms, however, it would clearly raise substantial revenue. For example, the Tax Policy Center estimates that the revenue-maximizing rate rises to 50% if stepped-up basis is repealed (Rubin, 2019). It would be even higher under reforms that tax gains (and allow deductions for losses) as they accrue, rather than waiting until they are realized.7 Nearly 40% of the wealth of the top 1% is in the form of accrued and unrealized capital gains. Moreover, the top 1% holds about half of all such unrealized gains.8

Taxing gains as they accrue is sometimes called a “mark-to-market” regime. Under mark-to-market, taxpayers would value all of their assets every year and either pay tax on the gain or deduct the loss. Such a system would eliminate the need for separate taxes on dividends and interest, since both would be considered part of any gain.

Given the difficulty of measuring the annual change in value of most privately held businesses and other illiquid assets, advocates of taxing gains as they accrue have generally proposed mark-to-market regimes only for publicly traded assets. But some advocate combining such an approach with a “retrospective” accrual regime for assets that are not publicly traded—which would impose tax only upon the sale of such assets but apply a deferral charge at the time of sale to minimize any benefit that had accrued from deferring tax payments on gains.

For example, suppose a wealthy investor purchases a resort for $100 million and it appreciates by $5 million each year for 10 years, at which point she sells it. Under a retrospective accrual tax, she would be taxed at the point of sale, but as if she was paying back taxes due, with interest, on her $5 million gain in each of the 10 years. Her tax liability would be higher than under our current realization-based system, which would also tax her on a $50 million gain, because of the interest charge.

This combined approach of mark-to-market for publicly traded assets and retrospective accrual taxation of all other assets has been proposed in conceptual form by Senator Ron Wyden (D-Oregon) and presidential candidate Julian Castro (Rubin, 2019; Wyden, 2019).

7 Another option is to apply higher capital gains rates as the amount of time the taxpayer holds an asset before realizing its accrued gains grows. We do not focus on this option here, but it would have similar effects.

8 Authors’ calculations from Federal Reserve Board (2017).
There is vast uncertainty in estimating the revenues that would be generated by either an accrual tax regime or a wealth tax. In Batchelder and Kamin (2019), we present some preliminary estimates of both approaches. Our estimates use the 2016 Survey of Consumer Finance, one of the best sources for wealth data, and use a range of tax avoidance assumptions to give some sense of the range of uncertainty.

We first estimate a proposal that applies mark-to-market to publicly traded assets, taxes the gains on such assets as ordinary income, and makes no changes to the taxation of gains on illiquid assets. We assume the top rate on ordinary income is 39.6% plus the SECA or NIIT tax of 3.8%. As summarized in Table 2, we estimate that this proposal would raise new revenue on the order of $1.7 trillion over 10 years if it were limited to roughly the top 1% (excluding additional income from the mark-to-market system, not total income, under about $100,000), and assuming a tax avoidance rate of 15%. For all of these estimates, we also provide figures in Table 2 assuming no avoidance and 30% avoidance to give a sense of the range of possible outcomes. However, publicly traded assets represent only about one-fifth of assets held by the top 1%, excluding retirement accounts and tax-exempt debt. Further, this estimate assumes there would be no change in the percentage of assets that are publicly traded, but such a regime would create vast incentives to privatize businesses and invest in other exempt assets.

Thus, we think the better approach is to apply an accrual tax to all assets but implement it only on a retrospective basis for assets that are not publicly traded. There are a number of different ways to do this, but under all of these approaches, gains on illiquid assets would only be taxed when the asset is sold. Importantly, such a retrospective regime should also treat gifts, bequests, and charitable contributions as a realization event in order to place illiquid assets on a similar footing as publicly traded assets taxed on a mark-to-market basis. Otherwise, significant tax avoidance opportunities would remain.

If a retrospective regime were applied to assets that are not publicly traded for the top 1% and also taxed gains on such assets as ordinary income, we estimate it would raise an additional $400 billion over 10 years, assuming a 15% avoidance rate. (See Batchelder and Kamin (2019) for detailed assumptions behind this estimate.)

Overall, we estimate an accrual tax would raise on the order of $2.1 trillion over 10 years if limited to the top 1% and assuming a 15% avoidance rate. It would raise

---

9 Authors’ calculations based on 2016 Survey of Consumer Finance (SCF).
10 As we describe in Batchelder and Kamin (2019), our estimate for applying a retrospective accrual tax to illiquid assets also very conservatively assumes that the only revenue raised from that regime is from taxing accrued gains on illiquid assets at death.
roughly $750 billion if limited to the top 0.1% instead.¹¹ We should emphasize that, unlike all the other revenue estimates in this paper, these estimates and those for a wealth tax assume no behavioral response, other than that embodied in the assumed tax avoidance rate.

One key issue in estimating revenue from these proposals is how pre-enactment, built-in gains on existing assets are treated in transition. When it comes to publicly traded assets, one approach would be to impose tax at the time the regime is implemented but allow payment of that liability over some period of time. Alternatively, payment on these pre-enactment, built-in gains could wait until sale of the asset or death, whichever is earlier. Our estimates assume something along the lines of the latter regime, by adopting the very conservative assumption that all pre-enactment, built-in gains on publicly traded assets are not marked-to-market until death. Most transition rules would likely raise more within the 10-year window,¹² and we hope to explore this further in future work.

4.2.1 Advantages

There are many advantages to a system that combines mark-to-market for publicly traded assets with retrospective accrual taxation for all other assets. It would raise a large amount of revenue almost exclusively from the wealthy. It would largely eliminate the ability to reduce tax liability by changing the timing of the sale of property. And it would effectively (under mark-to-market) or actually (under retrospective accrual) repeal stepped-up basis and eliminate the ability to avoid tax on gains by donating property to charities. Further, this increase in capital taxation could not be avoided through the kinds of tax planning maneuvers that allow multinational businesses to report a large share of their profits in tax havens. Accrual taxes would be imposed at the individual level on the multinational’s share price, which incorporates both domestic and foreign profits, and they would apply to all U.S. citizens, regardless of where they live. As a result, the location of profits reported by multinational enterprises—and the residence of the multinationals themselves—would be irrelevant for U.S. tax purposes.

¹¹ Some have suggested applying a retrospective accrual regime to both publicly traded and non-publicly traded assets (Grubert & Altshuler, 2016; Shakow, 1986). This would ensure that gains on publicly traded and non-publicly traded assets were taxed identically. This would certainly be an improvement over the current system, but it would not eliminate one reason that asset holders might defer gains: waiting for a reduction in rates or repeal of the retrospective regime. As a result, we view a combined system as a better approach. There is also a question of whether and how such regimes should be integrated with corporate income taxes. If gains and dividends on corporate stock are taxed at the individual level at ordinary rates and with no benefit to deferral, there is a logic to providing a credit at the individual level for taxes the corporation paid on its income (Grubert & Altshuler, 2016; Toder & Viard, 2016b). However, that logic breaks down if the mark-to-market regime is applied only to the very top of the income distribution, rather than more comprehensively. Thus, there may be a trade-off between comprehensive reform that integrates corporate- and individual-level taxes and limiting tax increases at the individual level to the very top.

¹² Specifically, we incorporate the effect of the built-in gain through existing estimates of the revenue raised from taxing accrued gains at death over the next decade. This is also our approach to the revenue from taxation of illiquid assets. See Batchelder and Kamin (2019).
For all these reasons, the revenue-maximizing capital gains rate would increase dramatically under an accrual tax system. Policymakers could then generate substantial revenues by increasing capital gains rates. By reducing or eliminating differences in the effective tax rates on ordinary income, capital gains, and dividends on a present value basis, policymakers could in turn eliminate or reduce many of the largest tax planning opportunities within our current system. Tax avoidance, with its accompanying fairness and efficiency costs, would decline.

The proposal would also be highly countercyclical, increasing the extent to which the federal fiscal system automatically stabilizes the macroeconomy. This is because accrual tax revenues and liabilities would more closely follow annual returns in the financial markets—swinging more widely from year-to-year—than under the current realization-based system.

Finally, there is substantial precedent in the United States for taxing gains as they accrue. Our current system taxes some securities (e.g., straddles) on a mark-to-market basis, and applies a retrospective accrual tax to some passive income earned in foreign corporations held by U.S. residents (so-called PFICs). Even more notably, we effectively apply an accrual tax approach to debt instruments through the original issue discount rules.

4.2.2 Potential Challenges

There are, however, a number of challenges associated with an accrual tax, some of which are substantial.

First, there would be additional administrative and compliance costs involved in reporting income on publicly traded assets on an annual basis based on changes in market values. That said, automated reporting by financial institutions could shield investors from much of this complexity.

Second, the heightened volatility of revenues under a mark-to-market regime is a double-edged sword. On the one hand, it would increase the extent to which federal fiscal policy automatically stabilizes the macroeconomy. On the other hand, if state governments also adopt the same regime (and many do piggyback off the federal tax system), it could increase the extent to which state policy magnifies economic cycles, by forcing states to cut spending during recessions in order to comply with their balanced budget rules. However, Toder and Viard (2016b) show this concern could be largely addressed by averaging the tax due over time. In addition, some states would receive an offsetting benefit: Accrual taxation would reduce the tendency of taxpayers in high-tax states to change their residence shortly before realizing large gains.

Third, restricting such a regime to the very wealthy would be relatively complicated compared to the wealth tax discussed next. One option would be to apply the regime universally and adjust tax rates to offset any undesired tax increase on average for
those below the very top. This would simplify the regime and allow relatively easy integration with the corporate tax system as everyone could receive credits for any corporate income taxes paid. But there would be no way to hold all those below a certain threshold harmless under this approach. While one could make sure that those in, for example, the bottom 99% do not face a tax increase on average (or even receive a tax cut), those with more capital holdings within this group would still tend to face tax increases. In addition, even if no one in this group faced a tax increase, they still might object to the complexity of complying with an accrual tax system.

Another option would be for policymakers to exempt taxpayers below a certain income or wealth threshold. But it is unclear how to treat taxpayers once they exceed the selected threshold. If taxpayers were then fully and permanently subject to the accrual tax, this cliff would create enormous incentives to stay below the threshold, potentially generating large economic distortions. Alternatively, policymakers could use some method to phase in the effects of the accrual tax regime.13

Fourth, the retrospective component of a partially retrospective accrual tax would necessarily be imprecise. Gains on assets do not accrue at a constant rate over time. Any deferral charge would necessarily be a rough approximation of the actual value of deferral to a specific taxpayer. Tax rates also change over time. Thus, while the retrospective component of the system would address valuation and liquidity concerns regarding illiquid assets, it would maintain some existing tax avoidance opportunities, while also introducing some new ones.14 But we should emphasize that these tax avoidance incentives, while meaningful, would generally be far smaller than under our current, realization-based system.

Finally, while there is substantial precedent in the United States for applying an accrual tax to some assets, no country taxes all assets on an accrual basis, even if restricted to the rich. Any time a new approach to taxation is enacted for a larger group of assets, there are inevitably unforeseen difficulties and unintended avoidance opportunities that can only be addressed gradually over time.

4.3 Wealth Tax

Another option for taxing the wealthy is to implement a new tax on wealth that is separate from our federal income, payroll, and wealth transfer tax systems.

13 See Batchelder and Kamin (2019) for a discussion of the trade-offs of each of these approaches.

14 For example, taxpayers holding assets that initially appreciate rapidly and then appreciate more slowly would have an incentive to hold such assets so that the appreciation was deemed to occur more gradually over time. As under current law, taxpayers who expect rates to fall would be incentivized to hold in order to take advantage of a future relatively low rate (Kamin & Oh, 2019; Hemel, 2019). Taxpayers with access to high-return investments might prefer the retrospective treatment to mark-to-market, and therefore would have an incentive to invest in privately held firms, not those that are publicly traded.
For instance, Senator Elizabeth Warren (D-Massachusetts) has proposed a 2% annual tax on net worth over $50 million—or the top 0.1%—and a 3% tax on net worth over $1 billion (Warren, 2019). The tax would apply to both domestic and foreign assets of U.S. citizens. To address incentives to expatriate, the proposal would also substantially increase the U.S. exit tax on Americans renouncing their citizenship. Senator Bernie Sanders (I-Vermont) has proposed a wealth tax with rates starting at 1% on net worth over $32 million and rising to 8% on net worth over $10 billion (Sanders, 2019).

There is considerable debate as to just how much revenue would be raised from a wealth tax. Saez and Zucman (2019a) estimate Warren's proposal would raise $2.75 trillion over 10 years (from 2019–2028), assuming a 15% avoidance rate. For purposes of consistency with our mark-to-market estimates, we have done a similar calculation using only the Survey of Consumer Finance (they average it with another data source) and assume the same 15% avoidance rate. Under these assumptions, we estimate the Warren proposal would raise about $2.6 trillion over 10 years (from 2021–2030). If a wealth tax instead was 2% and limited to the top 1% of wealth holders (net worth over about $10 million), we estimate it would raise about $5.2 trillion over 10 years, again assuming a 15% avoidance rate. As with the accrual tax (and unlike the other revenue estimates in this paper), these estimates assume no behavioral response other than that embodied in the tax avoidance rate.

Some experts have criticized the Saez and Zucman (2019) estimate as too high, taking issue with their data sources, methodology, or judgments about how Warren’s proposal would change as it made its way through the legislative and regulatory process (e.g., Smith, Zidar, & Zwick, 2019; Summers & Sarin, 2019; Sarin & Summers, 2019b). We discuss this ongoing debate in more detail in Batchelder and Kamin (2019). In light of this debate, and recognizing that there is substantial uncertainty about tax avoidance responses, we provide estimates assuming higher and lower tax avoidance rates. For instance, as summarized in Table 2, if there were 30% avoidance, the Warren proposal would raise approximately $2.0 trillion from 2021–2030.

We should also emphasize that considerable uncertainty exists regarding the total magnitude of wealth in the United States. This is not just a question of avoidance, but of what the levels of wealth are before such avoidance takes place. Different sources tend to show different levels and composition of wealth (e.g., Kopczuk, 2015; Bricker, Krimmel, Henriques, & Sabelhous, 2016; Saez & Zucman, 2016; Saez & Zucman, 2019c). Future research may shed additional light, as would of course the actual experience of a wealth tax in the United States if it were enacted and enforced.

4.3.1 Advantages

There are a number of advantages to a wealth tax on the most affluent. Wealth taxes can raise a large amount of revenue almost exclusively from the wealthy. Saez and Zucman (2019b) estimate that all of the revenue raised by Warren’s proposal would
be paid by the top 0.1% of households ranked by wealth. Ranked by income, 97% of the revenue would be paid by the top 1%. As with the personal income tax, relatively little of the burden should be shifted to other taxpayers, in part because the tax is based directly on the taxpayer’s wealth.

It is far easier to administer and comply with an exemption from a wealth tax than from an accrual tax. Wealth below the exemption is simply not taxed. Under an accrual tax (or at least one that avoids cliff effects), gains below the exemption are taxed on a realization basis, meaning the wealthy would probably need to comply with two different regimes with respect to each asset held.

Relative to raising ordinary rates and an FTT, a wealth tax would reduce deferral and lock-in incentives. While it would not change deferral incentives under the income tax, it would add an element of taxing capital that is not realization-based. As such, it could not be avoided by simply deferring gain and holding on to underperforming assets.

Like an accrual tax, a wealth tax could not be avoided through multinational businesses that shift reported profits (or actual economic activity) to low-tax foreign jurisdictions because it would effectively apply to the foreign profits of (U.S.- and foreign-resident) multinationals held by U.S. citizens.

In addition, a wealth tax may have a broader base than the alternatives, reducing tax avoidance opportunities and efficiency costs. It would definitely have a broader base than a financial transactions tax, which is limited to financial assets. But the relative breadth of its base compared to an accrual tax is largely a political economy question. Arguably it would be easier to include some assets in a wealth tax base. Theoretically, an accrual tax could apply to qualified retirement accounts, tax-exempt bonds, primary residences, and charitable transfers over which the donor retains some control. But this would be very challenging politically under an accrual tax because it builds on the income tax system, which currently exempts all or most returns on such assets from taxation. While including such assets in the base of a wealth tax would also be politically challenging, it might be somewhat easier because a wealth tax would be writing on a blank slate. With that said, these three categories comprise less than one-fifth of the wealth of the top 1% according to the Survey of Consumer Finance.15

If it is correct that it would be easier politically to apply a broad base to a wealth tax than an accrual tax, this would be a significant advantage. Either approach can result in extensive gaming if certain categories of assets are carved out or treated preferentially. For example, the Spanish wealth tax exempted some forms of closely held businesses and, over a short period of time, the exempted stock as a share of all closely held business stock grew from 15% to 77% (Alvaredo & Saez, 2009). This

---

15 Authors’ calculations based on Federal Reserve Board (2017).
relates to a further advantage: A number of other countries have enacted wealth taxes, providing precedents from which the United States could learn.

Finally, adding wealth as a separate tax base would arguably increase the fairness of the tax system as a whole. Tax fairness depends in part on how accurately the system distributes tax burdens based on how well-off taxpayers are. This raises the question of what is the best measure of being “well-off.” Income and consumption are excellent measures, but need not be the only ones. Wealth may provide additional information about well-being if, for example, it independently provides insurance against risks, access to better information, or political power. Empirically, wealth is also a powerful indicator of advantage, including inherited advantage, controlling for a variety of factors such as income and earnings (e.g., Hajat, Kaufman, Rose, Siddiqi, & Thomas, 2011; Jez, 2014; Boserup, Kopczuk, & Kreiner, 2016; Hotz, Wiemers, Rasmussen, & Koegel, 2018).

Despite these large, independent effects of wealth on well-being, our current tax system is regressive when measured by wealth, at least when one excludes human capital. Saez and Zucman (2019a) estimate that the bottom 99% of households pay about 7.2% of their wealth in federal, state, and local taxes, while the top 0.1% pay only 3.2%.

An accrual tax would reduce the regressivity of the current tax system by wealth, but not as effectively as a wealth tax. Accrual taxes apply heavier taxes to individuals whose wealth is appreciating rapidly, for example, entrepreneurs. They tax more lightly those whose wealth is growing slowly, such an heir to a large fortune who invests their portfolio conservatively.

While the U.S. wealth transfer taxes do tax wealth, they do not sufficiently address these fairness concerns either. Wealth transfer taxes are imposed only once per generation and do not apply to a large share of wealth that arguably should be counted when measuring relative affluence. Examples include wealth consumed during life or eventually given to family foundations or donor advised funds over which the donor maintains significant control. Wealth transfer taxes also do not apply to wealth transferred to heirs in exempt forms, such as paying for the private education of one’s descendants perpetually. In addition, even though the burden of wealth transfer taxes largely falls on the heirs of large fortunes and not decedents, they only partially correct for the fact that inherited income is tax-exempt under the income tax (Batchelder & Khitratakun, 2008). We tend to think U.S. wealth transfer taxes tax should be significantly strengthened and potentially replaced with an inheritance tax as proposed by Batchelder (2009). But even if it were, important arguments for a wealth tax would remain.

Furthermore (and relevant to debates about how much a wealth tax would raise), estate taxes are inherently more prone to avoidance than wealth taxes because they apply only at one point in time per generation. A variety of estate tax avoidance
strategies involve temporarily and artificially deflating the value of transferred assets at the point in time that the wealth transfer is deemed to occur—and therefore valued—for tax purposes (for a more detailed discussion and reform proposals, see Dodge, 2016; Batchelder & Kamin, 2019). The wealthy should be far less inclined to engage in such strategies under a wealth tax because doing so would entail ongoing—not temporary—restrictions on their powers over, and access to, their assets. In addition, assuming their heirs are also wealthy, any temporarily undervalued gifts and bequests would quickly be included at their correct value in the wealth tax base. The estate and gift taxes could only reach such temporarily undervalued wealth much later—if and when the heirs eventually transfer their inheritances to their children.

4.3.2 Potential Challenges

Despite these advantages, there are a number of potential objections to, or challenges associated with, a wealth tax.

Some object that a well-functioning income tax is a more efficient and fair way to tax the rich. Unlike an accrual tax and eliminating stepped-up basis, a wealth tax would not raise the revenue-maximizing capital gains rate. As a result, it would not eliminate barriers to equalizing the ordinary and capital gains rates, with all the attendant benefits of reducing tax avoidance and thereby increasing fairness and efficiency.

In addition, going forward, a wealth tax imposes a greater effective burden on the “normal” return to capital and less on rents (e.g., Kopczuk & Schrager, 2014). For instance, imagine two individuals save $100 million but one earns a 5% “normal” return and the other earns a 15% return—with 10 percentage points of that return reflecting “rents.” Under a 2% wealth tax (and ignoring any exemption), each would pay $2 million in taxes. The implicit income tax rate on the “normal” return would be 40%, while the implicit income tax rate on the “rents” would be zero. By contrast, an income tax could be set at a 20% rate to generate the same revenue, since it would also tax the above-market rate of return. It would tax the “normal” return and the “rents” at the same rate. In this way, an income tax would impose a greater burden on rents and less of a burden on the normal return as compared to a wealth tax generating the same revenue.

In some respects, the concern that a wealth tax would tax “normal” returns and rents at different rates is simply a way of saying that one believes income is a better measure of well-being than wealth. But there is another concern. If, as some economists believe, saving and investment decisions depend on the after-tax “normal” rate of return and not rents, a wealth tax would tend to generate greater distortions to such decisions than an equivalent income tax (e.g., U.S. Department of Treasury, 1977; President’s Advisory Panel on Tax Reform, 2005). But there is mixed evidence on the extent to which aggregate savings and investment is influenced by taxes at all (e.g., Elmendorf, 1996; Laitner & Juster, 1996; Dynan, Skinner, & Zeldes, 2002; Kopczuk
& Lupton, 2007) and, if so, whether it responds more to taxes on “normal” returns or rents (e.g., Batchelder, 2017b). Moreover, a wealth tax may encourage people to deploy their capital more productively by taxing low return assets at the same rate as those earning high returns (Guvenen et al., 2019).

A second potential concern with a wealth tax is liquidity. Because a wealth tax applies regardless of whether one’s assets are liquid or producing any income in the current year, it could create serious challenges if applied to middle-income households. This is a frequent objection to state-level property taxes and has driven many of the preferences and exemptions for certain categories of assets in the wealth taxes of other countries. But, if limited to the very wealthy, this concern has much less force. The wealthy can borrow against assets relatively easily and quickly. One oft-cited example is the $10 billion line of credit obtained by Oracle CEO Larry Ellison in 2014 (Thornton & Hendricks, 2019). It is possible that some minority owners of early-stage businesses could face liquidity challenges. These challenges would be heightened if the business obtains a very high valuation initially and subsequently fails. But by the time such taxpayers are worth $50 million, this seems unlikely. Moreover, a wealth tax could permit taxpayers to defer paying any tax due for several years with interest, as Warren has proposed. It could also allow taxpayers to average their wealth over several years to address situations where a taxpayer’s net worth briefly exceeds the threshold before returning to a level well below it.

Relative to the alternatives discussed thus far, a more serious drawback is that a wealth tax would create significant valuation challenges. According to the best available data, private businesses comprise up to half of the holdings of individuals whose net worth exceeds $50 million (IRS, 2018; authors’ calculations based on 2016 Survey of Consumer Finances, 2017). Wealthy individuals and the Internal Revenue Service (IRS) already often have to value private businesses and other hard-to-value assets without a market transaction. In the tax context, they do so for estate and gift taxes, or when claiming the charitable contribution deduction. In non-tax contexts, many large private businesses are valued on secondary markets (though frequently at a discount), and as part of mergers and acquisitions, obtaining venture capital funding, or issuing shares. Smaller businesses and assets like art are often valued as part of divorces, bankruptcies, or obtaining loans or insurance. But a wealth tax would require such valuations far more frequently. This could result in substantial tax avoidance, given the greater resources the wealthy can devote to valuation experts and litigation than the IRS. It could also create a large incentive to invest in private businesses, potentially reducing market transparency and liquidity.

The experience of other countries and recent empirical work provides grounds for hope that these valuation challenges could be effectively addressed. Several other countries use rules of thumb for valuing private businesses, such as the book value of assets plus a multiple of profits or sales (McDonnell, 2013; OECD, 2018). Smith, Zidar, and Zwick (2019) have recently developed and applied a detailed, industry-specific formula for estimating the value of private businesses held by the wealthy in
the United States. Wealth tax legislation or regulations could require these valuation formulas, or offer them as a safe harbor, while allowing taxpayers to prove a different value. Gamage (2019) supports relying exclusively on requiring valuation formulas in most cases in order to limit gaming. But it is also possible that any such formulas would not put private and publicly traded businesses on an equal footing once political economy considerations are taken into account.

This raises a more general concern: the potential for tax avoidance and evasion under a wealth tax. The number of OECD countries with a wealth tax has declined substantially over time, from 12 in 1990 to six today (Bunn, 2019; OECD, 2018). Some attribute the repeal of wealth taxes in these jurisdictions to excessive avoidance and evasion; others to their relatively narrow tax bases, which made them not worth the costs of administration; and still others to their relatively low exemptions, which generated political opposition (Saez & Zucman, 2019c; OECD 2018; Viard, 2019). Some wealth tax avoidance techniques would not transfer to the U.S. context. For example, some taxpayers avoided European wealth taxes by moving to other countries (Kleven, Landais, & Saez, 2013). But the United States, unlike every country that has implemented a wealth tax, taxes its citizens regardless of where they reside. The only way Americans can escape U.S. taxation is by giving up their U.S. citizenship, and even then the United States imposes a stiff exit tax, which Warren proposes to increase.

Relative to other options for raising a comparable amount of revenue from the very wealthy, it is unclear whether a wealth tax would entail more severe tax avoidance and evasion. The repealed European wealth taxes included a variety of exemptions for specific categories of assets, which facilitated avoidance and evasion, sometimes dramatically (OECD, 2018; Leiserson, McGrew, & Kopparam, 2019). There is a real risk that the United States would enact similar, asset-based exemptions and preferences as a wealth tax made its way through the political process.

On the other hand, as discussed, the risk of such exemptions and preferences may be lower under a wealth tax than under options using existing tax instruments as a matter of political economy. The U.S. income and wealth transfer taxes already have extensive and well-entrenched preferences for certain types of assets. By writing on a blank slate, a wealth tax might be able to avoid such preferences and reach forms of wealth that the United States has traditionally found politically challenging to tax, such as private foundations over which the donor maintains control.

Effectively enforcing a wealth tax would require substantial new enforcement resources for the IRS, and an expansion to our information reporting agreements with other countries. The United States already receives information on the foreign financial accounts of U.S. citizens in 113 countries under the FATCA regime and its successors (U.S. Department of the Treasury, 2019a). But the IRS currently lacks the resources to effectively use this data. While our existing information exchange agreements largely focus on financial assets, the OECD-led Common Reporting
Standard (CRS) covers non-financial assets, including trusts, and has over 100 signatories. The United States has not signed on to the CRS, but we nevertheless obtain information reported under it from other countries (Schneidman, 2019; OECD, 2019b).

Finally, a wealth tax could be struck down as unconstitutional on the grounds that it is a “direct tax,” which must be apportioned among the states on the basis of population under Article I, Section 9. We think a wealth tax is not a “direct tax” as a matter of law, and should therefore be upheld as constitutional (see, for example, Johnsen & Dellinger, 2018; Ackerman et al., 2019; Johnsen et al., 2019; Feldman, 2019; for a contrary view see Freeman, 2019; Khan, 2019). But the Supreme Court as currently constituted may nevertheless disagree.

What is clear is that any legal risk associated with enacting a wealth tax could be reduced if it were understood as a refinement to the income tax. For example, a wealth tax could be understood as a tax on imputed income from wealth (Cunningham & Schenk, 1992; Schenk, 2000; Gamage, 2019). This would follow the model of the Dutch dual income tax, which taxes the capital income of some assets based on an imputed return, not realized income (Cnossen & Bovenberg, 2001), and the former Columbian wealth tax, which was treated as a minimum income tax (Saez & Zucman, 2019c). Or it could be structured as an adjustment to marginal income tax rates based on wealth, much as we adjust marginal income tax rates based on family structure, age, the presence of capital income, and innumerable other factors (Glogower, 2019). Yet another possibility is to design a new tax that is a hybrid of an accrual and a wealth tax, perhaps using the wealth tax as a withholding mechanism or safe harbor under an accrual-based income tax.

While some suggest that the Supreme Court as currently constituted might also strike down a mark-to-market tax on capital gains, this seems far less likely. The constitutionality of the income tax is enshrined in the 16th Amendment. One Supreme Court case (Eisner v. Macomber, 1920), struck down application of the income tax where there was no realization. But it has been dramatically scaled back and essentially limited it to its facts. Virtually all commentators now agree the realization requirement is a mere administrative convenience and not constitutionally required (e.g., Hurley, 2008; Kornhauser, 2009; Toder & Viard, 2016a). Moreover, lower courts have declined to overturn several long-standing provisions that tax income on a mark-to-market basis, rather than when it is realized. Finally, any dubious arguments against the constitutionality of a mark-to-market tax do not apply to a retrospective accrual tax, which would only tax gains upon realization.

16 For examples, see Miller appendix in Toder & Viard, 2016a.
4.4 Financial Transactions Tax

A financial transactions tax (FTT) applies a tax to the sale of financial assets. An FTT is best viewed as a sales tax on securities. But it could also be viewed as combining elements of income and wealth taxation. Like the current income tax, an FTT is prompted by exchange of an asset. However, unlike the income tax, it is not imposed on the gain on the asset, but rather on the full value of the asset at that point—like in a wealth tax, though potentially multiple times per year. Unlike both, it is restricted to financial assets.

One FTT option outlined by the CBO is to apply a 10-basis-point (0.1%) tax to sales of stocks and debt obligations, and to payments made under derivative contracts (CBO, 2018). Transactions by foreigners on U.S. markets would be taxed, as would offshore trades by U.S. taxpayers. The tax would not apply to the initial issuance of stock or debt obligations, or to currency transactions or transactions involving short-term debt obligations. Extrapolating from CBO estimates, this option would raise about $810 billion over a decade (CBO, 2018). Like several of the other structural changes discussed above, revenue estimates of such a large-scale FTT are relatively uncertain and depend significantly on assumed effects on trading volume.

4.4.1 Advantages

As with the other progressive, structural changes discussed here, an FTT could raise substantial revenue primarily from the wealthy. However, an FTT also has some advantages as compared to these options.

Unlike any of them, a meaningful portion of the burden would fall on foreigners, which could be viewed as an advantage from a U.S. perspective. About 20% of U.S. long-term securities are held by foreign persons (U.S. Department of Treasury, 2019). Unlike a wealth tax, valuation is not a major barrier because the tax is imposed as the asset changes hands, often for cash. There should be no constitutional risk as the federal government’s power to tax transactions is well established. Unlike an accrual tax, an FTT seems relatively simple to understand.

There are already precedents for an FTT in the United States and other countries. The United States imposes a very small FTT to fund securities enforcement. Several other countries, including major trading centers like the United Kingdom and Hong Kong, impose much larger FTTs (Burman et al., 2016). These precedents provide lessons learned for the effective design of an FTT and reassurance that market disruptions would not be too severe.

Some also argue that an FTT would be a relatively efficient way to raise revenue from the wealthy (e.g., Baker, 2016). Overall, there is a compelling case that dynamics in the financial sector tend to lead to too much trading—trading where social costs exceed social benefits (Summers & Summers, 1989). One example is the extraordinarily large investments traders make in high-speed trading platforms and related infrastructure.
to beat out other traders in reacting to new information, all in pursuit of zero-sum gains (e.g., Budish, Cramton, & Shim, 2015; Baker & Gruley, 2019). Whether or not a broad FTT is the best response to these problems is a more difficult question. There are alternative tools that may more accurately target some of the significant failures in the markets for financial assets and reduce such wasteful behavior as high-frequency trading (Budish, Cramton, & Shim, 2015). But to the extent an FTT affects trading volume taking the form of these rent-seeking and speculative activities, it could curb the disruptive effects of such activities (e.g., the 2010 “flash crash”), while entailing relatively few efficiency costs.

4.4.2 Potential Challenges

Taking the other view, some are concerned that an FTT would have such large effects on trading volume that it would reduce liquidity, increase market volatility, and inhibit price discovery (Matheson, 2012; Habermeir & Kirlenko 2003).

An FTT is likely to reduce trading volume substantially (see Matheson, 2012 and Burman et al., 2016 for reviews). As a result, a broad FTT may impede some transactions whose benefits outweigh costs, but the key empirical question is how large such an effect would be (Matheson, 2011). This particular issue is not a concern for the other structural reforms discussed above because the present value of tax liability is not affected by the frequency of transactions. With that said, one estimate suggests that more than 50% of daily volume in the U.S. equities markets is driven by high-frequency traders (Meyer, Bullock, & Rennison, 2018). As a result, even a large decline in trading volume may not cut all that much into the “true” market liquidity that leads to price discovery for regular market participants.

Another drawback of an FTT is that the maximum amount it could raise is probably lower than the other options. Burman et al. (2016) estimate that an FTT would start losing revenue if the rate was over 0.34%, and this revenue-maximizing rate would raise only 17% more revenue than an FTT of 0.1%.

In addition, an FTT may not be as progressive as some of the other options, even though it would be highly progressive. There is some debate about whether the burden of an FTT would fall on all owners of capital by increasing financial asset prices with other asset prices adjusting, or just on the financial sector by reducing rents in that sector (for further discussion, see Burman et al., 2016; Baker, 2016). Either way, an FTT would differ from the other structural changes discussed because its statutory incidence would not fall exclusively on the wealthy, and its economic incidence might fall more on households below the top 99%. It would, however, still be highly progressive. For example, Burman et al. (2016) estimate that about two-thirds of the economic incidence would fall on the top 1% in the short run, and 40% on them in the long run. Once behavioral responses are taken into account, they argue the distributional effects would be even more progressive.
An FTT also could render the tax system less fair among the very wealthy. It would not directly burden very affluent individuals who trade their wealth rarely, if at all. For example, a billionaire whose wealth is almost exclusively held in stock of the company she founded wouldn’t owe any FTT on that wealth or its accrued gains until the point of sale, at which point there would be no deferral charge. Thus, her tax liability would be much lower than an individual with comparable income and wealth who trades her assets more frequently. A wealth or accrual tax would not entail this type of inequity.

Despite the models from other countries, there are several serious challenges in designing an FTT. In order to preserve liquidity, an FTT should probably include an exemption for market makers. Market makers are firms that stand ready to buy and sell a particular security on a regular and continuous basis at a publicly quoted price. But defining when a firm is acting as a market maker would be challenging.

Any FTT should also be designed not to drive up the prices on certain products too much because of cascading effects. For example, if an FTT applied to short-term Treasuries, it could inhibit their use for cash management because they have relatively low returns and are traded frequently. Again, determining where to draw the line on which securities should be exempt or eligible for lower rates would be difficult.

Further, the tax would have to be designed to address key avoidance techniques—including off-shoring of transactions and shifting across financial instruments. One key concern with an FTT is that it might drive transactions offshore. The tax should be designed to apply tax to any transaction involving a U.S. national (whether an entity or individual) and irrespective of whether the transaction occurs offshore. However, that would require the government collecting information on such offshore transactions, including when U.S. nationals disguise their transactions behind foreign corporations in which they are owners. One way to further expand the base to make it even harder to avoid an FTT would be to also apply it to any financial instrument issued by a U.S. company, irrespective of where it is traded and the nationality of the traders. The tax should also be applied across all types of securities to avoid shifting, but there may be no way to design an FTT that can’t be avoided at least to some degree by shifting across financial instruments.

A final and related drawback of an FTT stems from political economy considerations. An FTT would require highly technical rules and—to a greater extent than the other options—its burdens would be narrowly concentrated on a well-organized and highly resourced industry. This is a recipe for vociferous lobbying at both the legislative and regulatory stages (Mashaw, 1997; Kalaitzake, 2017). Absent sophisticated and well-resourced government actors and civil society groups, the net result could be a very watered down FTT that is easily avoided and raises relatively little revenue.
5. Conclusion

This paper has outlined policy options for raising a large amount of revenues primarily from the most affluent, including incremental approaches and four more structural changes: dramatically increasing the top tax rates on labor and other ordinary income; effectively taxing the wealthy on accrued gains as they arise and at ordinary rates; a wealth tax on high-net-worth individuals; and a financial transactions tax. It generally concludes that they all merit serious consideration and several are important complements to each other. For example, a dramatic increase in the top rates on labor and other income would function best if coupled with a partially retrospective accrual tax that taxes gains at higher rates. In practice, however, their relative strengths will turn to a large extent on how each is designed after it has made its way through the legislative and regulatory process.

References


LILY BATCHELDER

Robert C. Kopple Family Professor of Law, NYU School of Law

Lily Batchelder is the Robert C. Kopple Family Professor of Law at NYU School of Law and an affiliated professor at the NYU Wagner School of Public Service. Previously she served as Deputy Director of the White House National Economic Council and Deputy Assistant to the President under President Obama (2014-2015), and Majority Chief Tax Counsel for the U.S. Senate Committee on Finance (2010-2014). Batchelder’s scholarship and teaching focus on personal income taxes, business tax reform, wealth transfer taxes, retirement savings policy, and social insurance. She is on the board of Tax Analysts and the National Tax Association. Before joining NYU in 2005, Batchelder was an associate at Skadden, Arps, Slate, Meagher & Flom, director of community affairs for a New York state senator, and a client advocate for a social services organization in Ocean Hill-Brownsville, Brooklyn. She holds an AB from Stanford University, an MPP from the Harvard Kennedy School, and a JD from Yale Law School.

ERSKINE B. BOWLES

Co-Chair, Economic Strategy Group; President Emeritus, The University of North Carolina

Erskine B. Bowles is co-chair of the Aspen Economic Strategy Group and the chairman of the board of advisors of BDT Capital Partners. He began his business career at Morgan Stanley and went on to found and lead Bowles, Hollowell, Conner, which became the preeminent investment bank focused on middle market mergers and acquisitions. He subsequently co-founded the middle market private equity firm Carousel Capital, served as a senior advisor for Credit Suisse First Boston, and as a partner with the global buyout firm Forstmann Little & Co. He previously served on the boards of Morgan Stanley (Lead Director), First Union Corporation, Merck, VF, Cousins Properties (Lead Director), Norfolk Southern Corporation, General Motors, Belk and Facebook. In his public sector career, Erskine served in the Clinton Administration as SBA Administrator, Deputy White House chief of staff, and chief of staff. During his time in the White House, he negotiated the first balanced budget in a generation. Erskine co-chaired with Senator Alan Simpson the National Commission on Fiscal Responsibility and Reform, and from 2005-2011, he served as the president of the University of North Carolina.
WILLIAM G. GALE

Arjay and Frances Miller Chair in Federal Economic Policy in the Economic Studies Program, Brookings Institution

William G. Gale is the Arjay and Frances Miller Chair in Federal Economic Policy in the Economic Studies Program at the Brookings Institution. His research focuses on tax policy, fiscal policy, pensions and saving behavior. He is co-director of the Tax Policy Center, a joint venture of the Brookings Institution and the Urban Institute. He is also director of the Retirement Security Project. From 2006 to 2009, he served as vice president of Brookings and director of the Economic Studies Program. Gale is the author of Fiscal Therapy: Curing America’s Debt Addiction and Investing in the Future (Oxford University Press, 2019). He earned a Ph.D. in Economics from Stanford University and an A.B. from Duke University and attended the London School of Economics as an undergraduate.

AMY GANZ

Deputy Director, Economic Strategy Group

Amy Ganz is Deputy Director of the Aspen Economic Strategy Group. Prior to joining the Aspen Institute, she was a research associate at the NYU Furman Center, where she contributed to the center’s research on New York City housing, neighborhoods, and urban policy. Prior to NYU, she was a diversity partner at Andreessen Horowitz. She also served as Special Assistant to Special Olympics International Chairman and CEO Timothy Shriver. She holds a BA in economics and political science from Wheaton College and an MPA in public policy analysis from NYU Wagner School of Public Service.

CRAIG GARTHWAITE

Herman R. Smith Research Professor in Hospital and Health Service, Kellogg School of Management

Craig Garthwaite is the Herman R. Smith Research Professor in Hospital and Health Service, an Associate Professor of Strategy, and the Director of the Program on Healthcare at the Kellogg School of Management at Northwestern University. He is an applied microeconomist whose research examines the healthcare and biopharmaceutical sectors. His recent work focuses on questions related to social insurance programs and their effect on private firms and the provider market. Professor Garthwaite also studies questions of pricing and innovation in the biopharmaceutical sector. His research has appeared in journals such as the Quarterly Journal of Economics, the American Economic Review and the New England Journal of Medicine and has been profiled in media outlets such as the Wall Street Journal,
the Washington Post and Vox. He has also appeared as a guest on various television and radio shows such as Nightly Business Report and NPR Marketplace. Garthwaite received a B.A. and a Masters in Public Policy from the University of Michigan and his PhD in Economics from the University of Maryland. He has testified before the United States Senate and House of Representatives on matters related to competition in healthcare markets and pharmaceutical pricing.

**DAVID KAMIN**

*Professor of Law, NYU School of Law*

David Kamin is Professor of Law at NYU School of Law. His scholarship focuses on tax and budget policy, and he has published on issues ranging from tax planning to the tax code’s effect on inequality and poverty to the role of budget baselines in the legislative process. Before joining NYU Law in 2012, Kamin worked in President Obama’s administration. There, he served as Special Assistant to the President for Economic Policy and, before that, as Advisor to the Director of the Office of Management and Budget. In those roles, Kamin helped coordinate administration policy on federal tax and budget issues. Kamin earned a BA in economics and political science with highest honors from Swarthmore College in 2002, and he earned a JD magna cum laude from NYU Law in 2009.

**MELISSA S. KEARNEY**

*Director, Aspen Economic Strategy Group; Neil Moskowitz Professor of Economics, University of Maryland*

Melissa S. Kearney is the Neil Moskowitz Professor of Economics at the University of Maryland. Kearney’s research focuses on poverty, inequality, and social policy in the United States. Her work has been published in leading academic journals and frequently cited in the press. She is also Director of the Aspen Economic Strategy Group; a Research Associate at the National Bureau of Economic Research (NBER); and a non-resident Senior Fellow at Brookings. She is a senior editor of the Future of Children, an editorial board member of the American Economic Journal: Economic Policy, Journal of Economic Literature, and Demography, and she was previously co-editor of the Journal of Human Resources. She serves on advisory boards for the Notre Dame Wilson-Sheehan Lab for Economic Opportunities (LEO), Smith Richardson Foundation, and Russell Sage Foundation. Kearney previously served as Director of the Hamilton Project at Brookings and as co-chair of the MIT J-PAL State and Local Innovation Initiative. Kearney teaches Public Economics at both the undergraduate and PhD level at the University of Maryland. She holds a BA in Economics from Princeton University and a PhD in Economics from the Massachusetts Institute of Technology.
MAGNE MOGSTAD

Gary S. Becker Professor in Economics, University of Chicago

Magne Mogstad is the Gary S. Becker Professor in Economics at the University of Chicago and the Director of the Ronzetti Initiative for the Study of Labor Markets at the Becker Friedman Institute. His primary research fields are labor economics, public economics, and economic inequality, and his research focuses on how to address market failures and equalize opportunities. Magne has published extensively in leading scholarly journals, and he is currently an editor of the Journal of Political Economy. He earned his PhD in economics from the University of Oslo.

HENRY M. PAULSON, JR.

Co-Chair, Economic Strategy Group; Chairman, Paulson Institute

Hank Paulson is the founder and chairman of the Paulson Institute, which aims to foster a US-China relationship that maintains global order in a rapidly evolving world. He is also the co-chair of the Aspen Economic Strategy Group, and co-chair of the Bloomberg New Economy Forum Advisory Board. Paulson served as the 74th Secretary of the Treasury under President George W. Bush, from July 2006 to January 2009. Prior to that, he had a thirty-two year career at Goldman Sachs, serving as co-chairman and co-CEO beginning in 1998, and as chairman and CEO beginning in 1999. A lifelong conservationist, Paulson was Chairman of The Nature Conservancy Board of Directors and, prior to that, founded and co-chaired the organization’s Asia-Pacific Council. In 2011, he founded the Latin American Conservation Council, comprised of global business and political leaders, which he co-chaired until 2017. He also co-chaired the Risky Business Project from 2013-2017, a non-partisan initiative that quantified and publicized the economic risks of climate change in the United States. Earlier in his career, he was a member of the White House Domestic Council as well as a staff assistant at the Pentagon. Paulson is the author of three books, most recently Firefighting: The Financial Crisis and its Lessons (with Ben Bernanke and Tim Geithner). Other books include the bestsellers On the Brink and Dealing with China. He graduated from Dartmouth College and received an M.B.A. from Harvard University.
THOMAS PHILIPPON

Max L. Heine Professor of Finance, New York University, Stern School of Business

Thomas Philippon is the Max L. Heine Professor of Finance at New York University, Stern School of Business. Philippon was named one of the “top 25 economists under 45” by the IMF in 2014. He won the 2013 Bernácer Prize for Best European Economist under 40, the 2010 Michael Brennan & BlackRock Award, the 2009 Prize for Best Young French Economist, and the 2008 Brattle Prize for the best paper in Corporate Finance. He is an academic advisor to the Financial Stability Board and the Hong Kong Institute for Monetary and Financial Research. He was an advisor to the New York Federal Reserve Board and a board member of the French prudential regulatory authority from 2014 to 2019, and the senior economic advisor to the French finance minister in 2012-2013. Philippon graduated from Ecole Polytechnique, received a PhD in Economics from MIT and joined New York University in 2003.

NANCY L. ROSE

Head of the Department of Economics and Charles P. Kindleberger Professor of Applied Economics, Massachusetts Institute of Technology

Nancy L. Rose is the Head of the Department of Economics and Charles P. Kindleberger Professor of Applied Economics at the Massachusetts Institute of Technology, where her teaching and research focus on industrial organization, competition policy, and the economics of regulation. She served as Deputy Assistant Attorney General for Economic Analysis in the Antitrust Division of the U.S. Department of Justice from 2014 - 2016, and directed the National Bureau of Economic Research program in Industrial Organization from its creation in 1991 until her appointment to the Department of Justice. She is a member of the American Academy of Arts and Sciences and a Distinguished Fellow of the Industrial Organization Society. Her research includes analyses of economic regulation and firm behavior, labor rent-sharing and determinants of executive pay, and merger policy. Professor Rose received her Ph.D. in Economics from MIT and an A.B. in Economics and Government from Harvard University. Her accomplishments have been recognized by numerous fellowships over her career, including those from the Alfred P. Sloan Foundation, Center for Advanced Study in the Behavioral Sciences, Hoover Institution, John Simon Guggenheim Foundation, and National Science Foundation. Her professional service includes membership on the American Antitrust Institute Advisory Board, terms as Vice President and Executive Committee member of the American Economic Association (AEA), and independent directorships on both public and nonprofit boards.
JOHN VAN REENEN

Gordon Billard Professor of Management and Economics,
Massachusetts Institute for Technology

John Van Reenen has been the Gordon Billard Professor of Management and Economics at the Massachusetts Institute for Technology (jointly in the MIT Economics Department and Sloan Management School) since 2016. Prior to this, he was Director of the Centre of Economic Performance and Professor at the London School of Economics. He has published over a hundred papers with a particular focus on firm performance and the causes and consequences of innovation. He was the 2009 winner of the Yrjö Jahnsson Award (the European equivalent of the Clark Medal). He is a fellow of the British Academy and the Econometric Society and in 2017, was awarded an OBE for “services to public policy and economics” by the Queen. He has been a senior policy advisor to 10 Downing Street, the UK Secretary of State for Health and the European Commission. He received his BA from the University of Cambridge, his MSc from the London School of Economics and his PhD from University College London.

ALAN D. VIARD

Resident Scholar, American Enterprise Institute

Alan D. Viard is a resident scholar at the American Enterprise Institute (AEI), where he studies federal tax and budget policy. Prior to joining AEI, Viard was a senior economist at the Federal Reserve Bank of Dallas and an assistant professor of economics at Ohio State University. He has also been a visiting scholar at the US Department of the Treasury’s Office of Tax Analysis, a senior economist at the White House’s Council of Economic Advisers, and a staff economist at the Joint Committee on Taxation of the US Congress. Viard has taught public finance at Georgetown University’s Public Policy Institute and he co-hosted the New York University Law School tax policy colloquium in the spring 2015 semester. Earlier in his career, Viard spent time in Japan as a visiting scholar at Osaka University’s Institute of Social and Economic Research. Viard received his Ph.D. in economics from Harvard University and a B.A. in economics from Yale University.
## CO-CHAIRS

<table>
<thead>
<tr>
<th>Name</th>
<th>Title/Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Henry M. Paulson, Jr.</td>
<td>Founder and Chairman The Paulson Institute</td>
</tr>
<tr>
<td>Erskine Bowles</td>
<td>President Emeritus The University of North Carolina</td>
</tr>
<tr>
<td>James S. Crown</td>
<td>Chairman and CEO Henry Crown and Company</td>
</tr>
<tr>
<td>Robert K. Steel</td>
<td>Chairman Perella Weinberg Partners</td>
</tr>
</tbody>
</table>

## ASPEN INSTITUTE CHAIRMAN

<table>
<thead>
<tr>
<th>Name</th>
<th>Title/Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daniel Porterfield</td>
<td>President and CEO The Aspen Institute</td>
</tr>
</tbody>
</table>

## DIRECTOR

<table>
<thead>
<tr>
<th>Name</th>
<th>Title/Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melissa S. Kearney</td>
<td>Neil Moskowitz Professor of Economics The University of Maryland</td>
</tr>
<tr>
<td>Amy Ganz</td>
<td>The Aspen Institute</td>
</tr>
</tbody>
</table>

## MEMBERS

<table>
<thead>
<tr>
<th>Name</th>
<th>Title/Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wally Adeyemo</td>
<td>President Obama Foundation</td>
</tr>
<tr>
<td>Ben S. Bernanke</td>
<td>Distinguished Fellow in Residence Economic Studies Program Brookings Institution</td>
</tr>
<tr>
<td>Kelly Ayotte</td>
<td>Former U.S. Senator from New Hampshire</td>
</tr>
<tr>
<td>Joshua Bolten</td>
<td>President &amp; CEO Business Roundtable</td>
</tr>
<tr>
<td>Mary Barra</td>
<td>Chairman and CEO General Motors Company</td>
</tr>
<tr>
<td>Sylvia M. Burwell</td>
<td>President American University</td>
</tr>
</tbody>
</table>
Lanhee J. Chen  
David and Diane Steffy Fellow in American Public Policy Studies  
Hoover Institution  
Director of Domestic Policy Studies and Lecturer in the Public Policy Program  
Stanford University

Kenneth I. Chenault  
Retired CEO and Chairman  
American Express  
Chairman and Managing Director  
General Catalyst

Bob Corker  
Former U.S. Senator from Tennessee

David M. Cote  
Chairman and CEO  
GS Acquisition Holdings

James S. Crown  
Chairman and CEO  
Henry Crown and Company  
Chairman of the Board of Trustees  
The Aspen Institute

Ray Dalio  
Founder, Co-Chief Investment Officer, and Co-Chairman  
Bridgewater Associates, LP

Mitchell E. Daniels, Jr.  
President  
Purdue University

Brian Deese  
Global Head of Sustainable Investing  
BlackRock

Sue Desmond-Hellmann  
CEO  
Bill and Melinda Gates Foundation

Roger W. Ferguson, Jr.  
President and CEO  
TIAA

Laurence D. Fink  
Chairman and CEO  
BlackRock

Kristin Forbes  
Jerome and Dorothy Lemeson Professor of Management and Global Economics  
Sloan School of Management, Massachusetts Institute of Technology

Jason Furman  
Professor of the Practice of Economic Policy  
John F. Kennedy School of Government  
Harvard University

Timothy Geithner  
President  
Warburg Pincus LLC

Austan Goolsbee  
Robert P. Gwinn Professor of Economics  
Booth School of Business  
The University of Chicago

Danielle C. Gray  
SVP, Chief Legal Officer & Corporate Secretary  
Blue Cross and Blue Shield of North Carolina

Jon Gray  
President & Chief Operating Officer  
Blackstone

Reed Hastings  
Founder and CEO  
Netflix
Keith Hennessey  
Lecturer  
Graduate School of Business  
Stanford University

Marillyn Hewson  
Chairman, President, and CEO  
Lockheed Martin Corporation

Fred Hiatt  
Editorial Page Editor  
The Washington Post

Doug Holtz-Eakin  
President  
American Action Forum

Glenn Hubbard  
Dean Emeritus and Russell L. Carson  
Professor of Finance and Economics  
Columbia Business School

Greg Ip  
Chief Economics Commentator  
The Wall Street Journal

Neil Irwin  
Senior Economic Correspondent  
The New York Times

Joel Kaplan  
Vice President, Global Public Policy  
Facebook

Neel Kashkari  
President  
Federal Reserve Bank of Minneapolis

Melissa S. Kearney  
Neil Moskowitz Professor of Economics  
The University of Maryland

Ellen Kullman  
Former Chair and CEO  
DuPont

Edward P. Lazear  
Morris Arnold and Nona Jean Cox  
Senior Fellow  
Hoover Institution 
The Davies Family Professor of Economics  
Stanford Graduate School of Business

Maya MacGuineas  
President  
Committee for a Responsible Federal Budget

N. Gregory Mankiw  
Robert M. Beren Professor of Economics  
Harvard University

Magne Mogstad  
Gary S. Becker Professor in Economics and the College  
Kenneth C. Griffin Department of Economics  
University of Chicago

Marc H. Morial  
President and CEO  
National Urban League

Janet Murguía  
President and CEO  
UnidosUS

Phebe Novakovic  
Chairman and CEO  
General Dynamics

Michael A. Nutter  
David N. Dinkins Professor of Professional Practice in Urban and Public Affairs  
Columbia University School of International and Public Affairs

James Owens  
Chairman & CEO Emeritus  
Caterpillar Inc.
ABOUT THE ECONOMIC STRATEGY GROUP

The Economic Strategy Group (ESG), a non-profit program of the Aspen Institute, is composed of a diverse, bipartisan group of distinguished leaders and thinkers with the goal of promoting evidence-based solutions to significant U.S. economic challenges. Co-chaired by Henry M. Paulson, Jr. and Erskine Bowles, the ESG fosters the exchange of economic policy ideas and seeks to clarify the lines of debate on emerging economic issues while promoting bipartisan relationship-building among current and future generations of policy leaders in Washington.

CONTRIBUTORS

Lily Batchelder, New York University
William Gale, Brookings Institution
Craig Garthwaite, Northwestern University
David Kamin, New York University
Melissa S. Kearney, The University of Maryland
Thomas Philippon, New York University
Magne Mogstad, The University of Chicago
Nancy L. Rose, Massachusetts Institute of Technology
John Van Reenen, Massachusetts Institute of Technology
Alan D. Viard, American Enterprise Institute