BROOKINGS

www.brookings.edu

ISBN 978-0-8157-3946-3

Brookings Papers ON ECONOMIC ACTIVITY

BPEA
SPRING
2021

BROOKINGS

Janice Eberly and James H. Stock, Editors

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 2021

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 2021

JANICE EBERLY
JAMES H. STOCK
Editors

BROOKINGS INSTITUTION PRESS
Washington, D.C.

ISSN 0007-2303
ISBN: Print 978-0-8157-3946-3; ebook: 978-0-8157-3947-0

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 2021

BPEA LOOKING FORWARD TO THE NEXT FIFTY YEARS

Since its founding in 1970, *Brookings Papers on Economic Activity* (*BPEA*) has published high-quality independent research on timely economic and policy challenges. Currently led by coeditors Janice Eberly and James Stock, *BPEA* is widely recognized as a premier economics journal with a long-standing reputation for rigorous analysis and real-world application and policy relevance.

The year 2020 marked the fiftieth anniversary of *BPEA*. For over half a century, *BPEA* has attracted top talent in the field to serve as editors, authors, discussants, and advisers, including more than twenty Nobel laureates, Federal Reserve chairs, members of the President's Council of Economic Advisers, and chief economists from both public and private financial institutions. Three panel papers in this issue by longtime contributors to *BPEA* highlight *BPEA*'s seminal research over the years in areas at the heart of macroeconomic policymaking: labor markets, productivity and growth, and monetary policy.

BPEA addresses policy issues as they are emerging, such as the financial crisis in 2008, Brexit in 2016, and most recently, the COVID-19 pandemic in 2020 and 2021. *BPEA* convened two conferences in 2020 uniquely devoted to exploring the pandemic's impact on economic activity, financial markets, and individual lives. While the papers presented at those conferences represented the best thinking of economists early in the pandemic, three panel papers on COVID-19 in this issue investigate the pandemic's lasting economic effects and the efficacy of policy responses after one year.

This issue also features four papers that explore a wide range of important policy issues, including public pensions, education, financial regulation, and the criminal justice system. Ideas launched at *BPEA* often become policy soon afterward. In recent years, major findings have changed how we think about the student loan crisis, the high cost of health care, and long-term unemployment among American workers. As *BPEA* looks forward to the next fifty years, it continues to convene distinguished economic experts and act as a trusted platform for lively, intellectual debate and discussion. The need for this platform to generate and debate policy-relevant research is as important now—if not more so—than it was fifty years ago, and we are committed to ensuring that this enterprise thrives for the next fifty years.

PURPOSE The *Brookings Papers on Economic Activity* (*BPEA*) publishes research on current issues in macroeconomics, broadly defined. The journal emphasizes rigorous analysis that has an empirical orientation, takes real-world institutions seriously, and is relevant to economic policy. Papers are presented and discussed at conferences held twice each year, and the papers and discussant remarks from each conference are published in the journal several months later. Research findings are described in a clear and accessible style to maximize their impact on economic understanding and economic policymaking; the intended audience includes analysts from universities, governments, and businesses. Topics covered by the journal include fiscal and monetary policy, consumption and saving behavior, business investment, housing, asset pricing, labor markets, wage and price setting, business cycles, long-run economic growth, the distribution of income and wealth, international capital flows and exchange rates, international trade and development, and the macroeconomic implications of health care costs, energy supply and demand, environmental issues, and the education system.

We would like to thank the supporters of the *BPEA* conference and journal, including the Alfred P. Sloan Foundation; BlackRock Global Fixed Income; General Motors Company; the National Science Foundation, under grant no. 1756544; and State Farm Mutual Automobile Insurance Company. We gratefully acknowledge Dina Axelrad Perry for establishing the George L. Perry and the William C. Brainard *BPEA* Chair.

CALL FOR PAPERS Most papers that are presented at the *BPEA* conferences and appear later in the journal are solicited by the editors, but the editors also consider unsolicited proposals. Editorial decisions are typically made nine months prior to each conference—proposals received by December 1 are considered for the following fall conference, and those received by June 1 for the spring. However, qualified proposals may be considered for conferences outside of the normal evaluation timeline, depending on the timeliness of the topics and the program needs. Proposals from early career researchers, and members of underrepresented groups in the economics profession are encouraged. Proposals can be submitted at https://www.brookings.edu/bpea-for-authors/.

ACCESSING THE JOURNAL All past editions of *BPEA*, including versions of the figures in color—along with appendix materials, data, and programs used to generate results—are made freely available for download at www.brookings.edu/bpea/search. To purchase subscriptions or single copies, visit

www.brookings.edu/press, or contact the Brookings Institution Press at (866) 698-0010 or P.O. Box 465, Hanover, PA 17331-0465. Brookings periodicals are available online through both the Online Computer Library Center (contact the OCLC subscription department at (800) 848-5878) and Project Muse (http://muse.jhu.edu). Archived issues of *BPEA* are also available through JSTOR (www.jstor.org).

EDITORS, AUTHORS, DISCUSSANTS, ADVISERS, PARTICIPANTS, AND STAFF FOR THE ONE HUNDRED AND TWELFTH CONFERENCE

Henry Aaron *Brookings Institution*
Stephanie Aaronson *Brookings Institution*
Katharine Abraham *University of Maryland*
Viral Acharya *New York University Stern School of Business*
Şenay Ağca *National Science Foundation*
George Akerlof *Georgetown University*
Philippe Andrade *Federal Reserve Bank of Boston*
Andrew Atkeson *University of California, Los Angeles*
Martin Baily *Brookings Institution*
David Baqaee *University of California, Los Angeles*
Sandra Barbosu *Alfred P. Sloan Foundation*
Joe Beaulieu *Brevan Howard*
Ben Bernanke *Brookings Institution*
Sandra Black *Columbia University*
Olivier Blanchard *Peterson Institute for International Economics*
Alan S. Blinder *Princeton University*
Michael Boldin *Temple University*
Barry Bosworth *Brookings Institution*
William Brainard *Yale University*
Steven Braun *Council of Economic Advisers*
Jesse Bricker *Federal Reserve Board of Governors*
Ralph Bryant *Brookings Institution*
Elaine Buckberg *General Motors*
Gary Burtless *Brookings Institution*
Gerald Cohen *Mastercard*
Matthew Cook *Arnold Ventures*
Steven Davis *University of Chicago Booth School of Business*
Jason DeBacker *University of South Carolina*
Ryan Decker *Federal Reserve Board of Governors*
Brad Delong *University of California, Berkeley*
Will Dobbie *Harvard University*
Jane Dokko *Federal Reserve Bank of Chicago*
John Driscoll *Federal Reserve Board of Governors*
Janice Eberly *Northwestern University*
Wendy Edelberg *Brookings Institution*
Barry Eichengreen *University of California, Berkeley*
William English *Yale University*
Charles Evans *Federal Reserve Bank of Chicago*
John Fernald *Federal Reserve Bank of San Francisco/INSEAD*
Raquel Fernandez *New York University*
James Feyrer *Dartmouth College*
Kristin Forbes *MIT Sloan School of Management*
Benjamin Friedman *Harvard University*
Jason Furman *Harvard University*
Joseph Gagnon *Peterson Institute for International Economics*
William Gale *Urban-Brookings Tax Policy Center*

Amy Ganz *Aspen Institute*
Karen Glenn *Social Security Administration*
Austan Goolsbee *University of Chicago Booth School of Business*
Robert J. Gordon *Northwestern University*
Tracy Gordon *Urban-Brookings Tax Policy Center*
Egor Gornostay *Peterson Institute for International Economics*
Daniel Goroff *Alfred P. Sloan Foundation*
Steve Goss *Social Security Administration*
Joshua Gotbaum *Brookings Institution*
Carol Graham *Brookings Institution*
Robin Greenwood *Harvard Business School*
Erica Groshen *Cornell University*
Kwabena Gyimah-Brempong *National Science Foundation*
Robert E. Hall *Stanford University*
Peter Henry *New York University Stern School of Business*
Brad Hershbein *W. E. Upjohn Institute for Employment Research*
Bart Hobijn *Arizona State University, Tempe*
Douglas Holtz-Eakin *American Action Forum*
Harry Holzer *Georgetown University*
Caroline M. Hoxby *Stanford University*
Jennifer Hunt *Rutgers University*
Erik Hurst *University of Chicago Booth School of Business*
Brian A. Jacob *University of Michigan*
Lawrence Katz *Harvard University*
Anupam Khanna *NASSCOM*
Michael Kiley *Department of the Treasury*
Mervyn King *New York University Stern School of Business*
Aaron Klein *Brookings Institution*
Jeffrey Kling *Congressional Budget Office*
Donald Kohn *Brookings Institution*
Randy Kroszner *University of Chicago Booth School of Business*
Fabian Lange *McGill University*
Jamie Lenney *Bank of England*
Deborah Lucas *Massachusetts Institute of Technology*
Byron Lutz *Federal Reserve Board of Governors*
Nancy Lutz *National Science Foundation*
N. Gregory Mankiw *Harvard University*
Oscar Mendez *Alfred P. Sloan Foundation*
Karel Mertens *Federal Reserve Bank of Dallas*
Andrew Metrick *Yale University*
Laurence Meyer *LH Meyer*
Gian Maria Milesi-Ferretti *Brookings Institution*
Conrad Miller *University of California, Berkeley*
Frederic Mishkin *Columbia University*
Robert Moffitt *Johns Hopkins University*
Adele Morris *Brookings Institution*
Christopher Nekarda *Federal Reserve Board of Governors*
Maurice Obstfeld *University of California, Berkeley*
Sanjay Patnaik *Brookings Institution*
Karen Pence *Federal Reserve Board of Governors*
George Perry *Brookings Institution*
Thomas Philippon *New York University Stern School of Business*
Jonathan Pingle *BlackRock*
Richard Portes *London Business School*
Giorgio Primiceri *Northwestern University*
Joshua Rauh *Stanford Graduate School of Business*

Richard V. Reeves *Brookings Institution*
David Reifschneider *formerly of the Federal Reserve Board of Governors*
Ricardo Reis *London School of Economics and Political Science*
Giovanni Ricco *University of Warwick*
John Roberts *Federal Reserve Board of Governors*
Christina D. Romer *University of California, Berkeley*
David Romer *University of California, Berkeley*
Krista Ruffini *Georgetown University*
John Sabelhaus *Brookings Institution*
Ayşegül Şahin *University of Texas, Austin*
Claudia Sahm *Jain Family Institute*
Diane Schanzenbach *Northwestern University*
Finn Schüle *Brown University*
Jay Shambaugh *George Washington University*
Matthew Shapiro *University of Michigan*
Louise Sheiner *Brookings Institution*
Hyun Song Shin *Bank for International Settlements*
Sven Sinclair *Social Security Administration*
Tara Sinclair *George Washington University*
Anna Stansbury *Peterson Institute for International Economics*
Mark Steinmeyer *Smith Richardson Foundation*
James H. Stock *Harvard University*
Michael Strain *American Enterprise Institute*
Phillip Swagel *Congressional Budget Office*
Daniel Tarullo *Harvard University*
Robert Tetlow *Federal Reserve Board of Governors*
Jesse Thompson *Federal Reserve Bank of Dallas*
Sarah Turner *University of Virginia*
Andrés Velasco *London School of Economics and Political Science*
Stan Veuger *American Enterprise Institute*
Alan Viard *American Enterprise Institute*
Ivan Vidangos *Federal Reserve Board of Governors*
Polina Vlasenko *Social Security Administration*
Alice H. Volz *Federal Reserve Board of Governors*
William Wascher *Federal Reserve Board of Governors*
Mark Watson *Princeton University*
Tara Watson *Williams College*
David Wessel *Brookings Institution*
Johannes Wieland *University of California, San Diego*
David Wilcox *Peterson Institute for International Economics*
Jonathan Willis *Federal Reserve Bank of Atlanta*
Justin Wolfers *University of Michigan*
Abigail Wozniak *Federal Reserve Bank of Minneapolis*
Crystal Yang *Harvard University*
Luigi Zingales *University of Chicago Booth School of Business*
Eric Zitzewitz *Dartmouth College*

Francisca P. Alba *Brookings Institution*
Janina Broker *Brookings Institution*
Haowen Chen *Brookings Institution*
Siddhi Doshi *Brookings Institution*
Siobhan Drummond *Brookings Institution*
Swati Joshi *Brookings Institution*
Tyler Powell *Brookings Institution*
Ariel Gelrud Shiro *Brookings Institution*
David Skidmore *Brookings Institution*

JAMIE LENNEY
Bank of England

FINN SCHÜLE
Brown University

BYRON LUTZ
Federal Reserve Board of Governors

LOUISE SHEINER
Brookings Institution

The Sustainability of State and Local Pensions: A Public Finance Approach

ABSTRACT In this paper we explore the fiscal sustainability of US state and local government pension plans. Specifically, we examine whether, under current benefit and funding policies, state and local pension plans will ever become insolvent and if so, when. We then examine the fiscal cost of stabilizing pension debt as a share of the economy and examine the cost associated with delaying such stabilization into the future. We find that, despite the projected increase in the ratio of beneficiaries to workers as a result of population aging, state and local government pension benefit payments as a share of the economy are currently near their peak and will eventually decline significantly. This previously undocumented pattern reflects the significant reforms enacted by many plans which lower benefits for new hires and cost-of-living adjustments often set beneath the expected pace of inflation. Under low or moderate asset return assumptions, we find that few plans are likely to exhaust their assets over the next few decades. Nonetheless, under these asset returns, plans are currently not sustainable as pension debt is set to rise indefinitely; plans will therefore need to take action to reach sustainability. But the required fiscal adjustments are generally moderate in size and in all cases are substantially lower than the adjustments required under the typical full prefunding benchmark. We also find generally modest returns, if any, to starting this stabilization process now versus

Conflict of Interest Disclosure: The authors did not receive financial support from any firm or person for this paper or from any firm or person with a financial or political interest in this paper. They are currently not officers, directors, or board members of any organization with an interest in this paper. No outside party had the right to review this paper before publication. The analysis and conclusions reached in the paper are the authors' alone and do not indicate concurrence by the Bank of England or members of the Monetary Policy Committee, Financial Policy Committee, Prudential Regulation Authority Board, or the Board of Governors of the Federal Reserve.

a decade in the future. Of course, there is significant heterogeneity, with some plans requiring very large increases to stabilize their pension debt.

State and local government pension plans are important economic institutions in the United States. They hold nearly $5 trillion in assets; their annual benefit payments to beneficiaries are equal to about 1.5 percent of national GDP; over 11 million beneficiaries rely on these payments to sustain themselves in retirement. In recent years, attention has focused on the plans' large unfunded liabilities; recent estimates indicate that the obligations of public pension funds exceed their assets by around $4 trillion.[1]

The magnitude of these unfunded liabilities has generated widespread concern; indeed, public pensions are often viewed as being in a state of crisis, with the threat of default looming (online appendix figure A1).[2] But it has been understood at least since Samuelson (1958) that the existence of unfunded liabilities does not necessarily imply that a pension plan is unsustainable, in the sense that it will require outside funding to avoid default. Fully unfunded, pay-as-you-go (PAYGO) pension systems can be fiscally sustainable. Moreover, failure to prefund does not necessarily imply future fiscal costs, a corollary to the idea that public debt may have no fiscal cost in low interest rate environments (Blanchard 2019).

This paper focuses on state and local government pension systems as we find them today—that is, partially prefunded and therefore also partially pay-as-you-go—and asks if, under current policies and funding levels, state and local pension plans are fiscally sustainable over the medium and longer run and if not, what changes are needed? To answer this question, we project the annual cash flows of state and local pensions benefits. We find that pension benefit payments in the United States, as a share of the economy, are currently near their peak and will remain there for the next two decades. Thereafter, the reforms instituted by many plans will gradually cause benefit

1. See Rauh (2017) and "L.120 State and Local Government Employee Retirement Funds," Board of Governors of the Federal Reserve, https://www.federalreserve.gov/apps/fof/DisplayTable.aspx?t=l.120.

2. Commentary from academics include the claim that "the threat of default looms" for public pensions (Shoag and Farrell 2017, 4) and the statement that these pensions have failed to "provide economic security in old age in a financially sustainable way" (Novy-Marx and Rauh 2014b, 47). Members of Congress have expressed concern that state and local pensions are "unsustainable" and that requests for bailouts from the federal government are inevitable (US Congress Joint Economic Committee 2012). Finally, a major financial institution states that "there are no solutions for some plans given how underfunded they are" (Cembalest 2018, 2).

cash flows to decline significantly. This is a new and important finding in terms of the fiscal stability of these plans as it indicates that the cash flow pressures they currently face will eventually recede.

In terms of sustainability, we find that under low or moderate real asset return assumptions (0.5 percent and 2.5 percent) and a risk-free real discount rate (0.5 percent), for the United States as a whole, state and local pensions are not currently sustainable in the sense that pension debt as a share of the economy is set to rise indefinitely. That said, at the 2.5 percent asset return assumption, pension debt can generally be stabilized with only moderate fiscal adjustments—a conclusion which broadly holds across scenarios in which governments act to stabilize pension debt over the long run, medium run, and immediately. Under low asset returns, the required adjustments are generally larger, but are nonetheless much smaller than those required to achieve full funding over thirty years. Notably, there appear to be only modest returns to starting this stabilization process now versus a decade in the future: neither the level at which debt stabilizes as a share of the economy nor the contribution change needed to achieve stabilization increases significantly when the start of the stabilization process is pushed ten years out. Overall, while achieving fiscal stability will require adjustments, our results suggest there is no imminent crisis for most public pension plans.

Of course, there is significant heterogeneity across plans, with some plans requiring large contribution increases to achieve stability. That said, the plans that require the largest adjustments are not particularly those that are the least funded, reflecting the fact that our focus is on debt stabilization, not full funding. Of course, one might suspect that the least well-funded plans got that way by failing to make sufficient contributions and by ignoring looming imbalances. But we find that many of the most poorly funded plans have in recent years undertaken the largest reforms and increased contribution rates the most; in so doing, many of these poorly funded plans have already made significant progress toward stabilizing their pension debt.

Our focus on pension sustainability, as opposed to the more typical focus on a full prefunding benchmark, is useful and appropriate. First, it provides a clear answer to the pressing question of whether public pensions are likely to spark a fiscal crisis and when. Failure to fully prefund, in isolation, need not spark a crisis. Second, it is consistent with history; in aggregate, these plans have always operated far short of full prefunding. Third, full prefunding is not necessarily welfare enhancing, as we discuss below.

In terms of methodology, we reverse engineer the future stream of pension benefit payments using the method pioneered by Novy-Marx and Rauh

(2011) and also used in Lutz and Sheiner (2014). We use these projected cash flows, in conjunction with economic and demographic assumptions, to analyze the future evolution of each plan's pension debt. We employ this methodology on a sample of forty state and local pension systems that matches the national distribution of plans in terms of both mean and variance for multiple plan characteristics—for example, the funding ratio. For our main stabilization exercises, we use three deterministic rates of assets returns: the expected return and two lower rates of return that can be viewed as accounting for risk under a certainty-equivalent approach. One of these rates is a market-based risk-free rate of return. We also present an exercise in which realized asset returns are allowed to vary stochastically, allowing us to assess the full distribution of future pension debt and assets.

Our findings have significant policy relevance beyond directly addressing the sustainability of public pension plans. State and local governments have been ramping up pension plan contributions substantially in the years since the financial crisis (online appendix figure A2). These increased contributions come at a significant opportunity cost. Despite a long economic expansion prior to the COVID-19 pandemic, provision of the core public goods provided by these governments remained depressed: real per capita spending on infrastructure stood at about 25 percent below its previous peak, and state and local government employment per capita also remained well below its previous peak.[3] Notably, much of this relative decline in state and local government employment occurred in the K-12 and higher education sectors.[4] Thus, while pension contributions had been rising at a rapid clip, core investments in education and infrastructure were lagging. Finally, our results have important implications for intergenerational equity. If existing unfunded liabilities are fiscally sustainable, then concern for intergenerational equity may well dictate that they be paid off only very slowly, if at all, so as not to overly burden a single generation.

The remainder of the paper is structured as follows: section I provides background information, including a discussion of state and local pensions, PAYGO pension sustainability, public debt sustainability, and past research on state and local pension sustainability. Section II describes the data and sample selection, section III outlines our methodology, section IV presents the results on pension sustainability under current funding levels and benefit parameters, section V presents the results on the contribution changes required

3. Authors' calculation based on the Bureau of Economic Analysis, National Income and Product Accounts, tables 2.1 and 3.9.6.

4. Authors' calculations based on the Bureau of Labor Statistics, Current Employment Statistics (Establishment Survey).

to stabilize pension debt, section VI presents the results for the stochastic exercises, and section VII concludes.

I. Background

I.A. Pension Prefunding and Implicit Pension Debt Sustainability

In order to value implicit pension debt, a rate must be chosen with which to discount the future benefit payments. State and local governments have typically chosen to use a discount rate equal to the assumed rate of return on risky plan assets. However, standard financial principles of valuation suggest that a stream of future payments should be discounted at a rate which reflects the riskiness of the future stream of payment, which depends on the probability that the payments will be honored, among other factors. Given the relatively strong legal protections surrounding these payments, it is appropriate to use a discount rate lower than that implied by the expected return on the risky assets held by pension plans (Novy-Marx and Rauh 2011; Lucas 2012).[5] With lower discount rates, pension debt is typically much larger than stated in annual government accounting statements and most plans are far from being fully prefunded—that is, assets are well below the present value of future benefit payments (Novy-Marx and Rauh 2011).

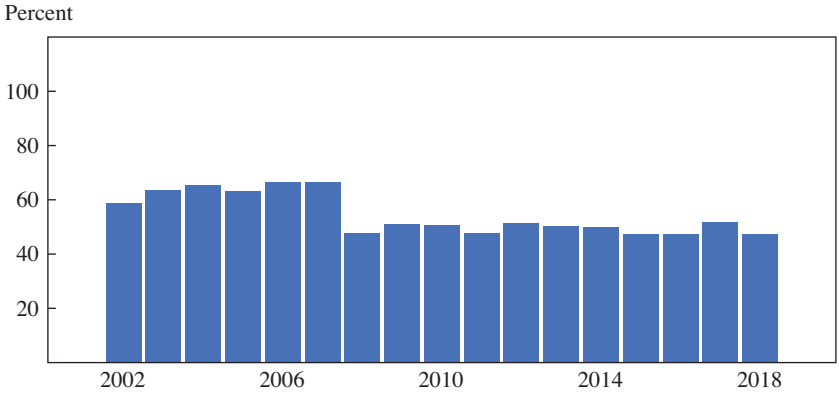
Figure 1 displays the aggregate funding ratio—the ratio of plan assets to the present discounted value of future obligations—for state and local government pensions from the Financial Accounts of the United States.[6] These estimates use the AAA corporate bond interest rate as the discount rate. Over roughly the last fifteen years, state and local pension plans have never exceeded 67 percent prefunding and averaged 55 percent prefunding.[7] Looking back as recently as 1978, one in six pension plans did not prefund to any degree, only 20 to 30 percent of plans were making sufficient contributions to prevent their unfunded liabilities from growing, and a quarter of local plans did not employ actuarial valuations and therefore could not even assess their funding level (United States Congress 1978). Thus, in aggregate, these plans have always operated well short of full prefunding.

5. The precise discount rate that should be used remains subject to debate, with some arguing for a risk-free rate (Brown and Wilcox 2009; Novy-Marx and Rauh 2009) and others arguing for a somewhat higher rate, such as that implied by state general obligation debt (CBO 2011) or a high-grade corporate bond yield (Lenze 2013; Lucas 2017).

6. "Financial Accounts of the United States—Z.1," Board of Governors of the Federal Reserve System, https://www.federalreserve.gov/releases/z1/.

7. Even using the plans' own elevated discount rates, these plans rarely have been fully prefunded, averaging just 83 percent funded over the past thirty years (see online appendix figure A3).

Figure 1. State and Local Government Pension Funding Ratios under AAA Corporate Bond Interest Rate



Source: Financial Accounts of the United States; see Hoops, Smith, and Stefanescu (2016) for methodology.

Moreover, the heavy emphasis on full prefunding in discussions of state and local pensions is a relatively recent development. As recently as 2008, many analysts considered a funding ratio of 80 percent to be sound practice (Government Accountability Office 2008).

It is often assumed that this failure to fully prefund the obligations is inappropriate or undesirable. For example, with regard to past academic work, Boyd and Yin (2016b) explicitly state that full prefunding is the proper goal for plans; in many other cases the position is taken more implicitly—for example, focusing analysis on the fiscal costs of transitioning to full funding (Novy-Marx and Rauh 2014b). With regard to policy makers, the nation's largest state and local pension plan explicitly advocates for full funding, stating that the "ideal level" of prefunding is 100 percent.[8] Along similar lines, the blue ribbon panel commissioned by the Society of Actuaries "wholeheartedly believes that . . . plans should be pre-funded" (Society of Actuaries 2014, 19). Finally, ratings agencies typically view "underfunding of pension . . . benefits as [a] key credit issue."[9]

Yet neither in terms of ex ante voter welfare or ongoing fiscal sustainability is the case for the full prefunding of public pensions clear (Brown, Clark, and Rauh 2011). In terms of fiscal sustainability, an unfunded PAYGO

8. From the California Public Employees' Retirement System, *Annual Review of Funding Levels and Risks*, 2014; no longer available online.

9. "U.S. Public Finance 2018 Year in Review," S&P Global Ratings.

pension system—such as the US Social Security system—can be fiscally sustainable in the sense that it requires no outside funding.[10] In particular, a fully unfunded PAYGO system can honor obligations without recourse to outside funding as long as the internal rate of return paid to beneficiaries does not exceed the growth rate of the wage base, equal to working-age population growth plus productivity growth (Samuelson 1958). Thus, these programs are only unsustainable if their costs rise at a faster pace than the underlying stream of revenue with which they are funded; such an event is typically caused by (1) demographic changes that increase the growth in outlays or lower the growth of revenues and (2) benefits rising faster than the underlying source of revenue because of increasing benefits promised over time. Mature, partially funded systems—which combine partial prefunding with partial PAYGO—can remain sustainable even in the face of adverse shocks, as accumulated assets provide a buffer. State and local pension plans almost always fall into this partially prefunded category.

More broadly, governments typically hold debt, and unfunded pension liabilities are simply a form of (implicit) debt. Such public debt can be sustainable as long as the government makes appropriate service payments on it. Pension debt stability when the growth rate of the economy and the interest rate are constant is illustrated by the following identity:

$$c_t = nc_t + \frac{(r-g)}{(1+g)}d \tag{1}$$

where c_t is the pension contribution as a share of the GDP required to keep the share of implicit pension debt to GDP ($d = d_t = d_{t-1}$) stable; nc_t is the normal cost—the liability accrued in period t for current employees' future pension benefits—as a share of GDP; g is the rate of GDP growth; and r is the interest rate.[11] When the rate of interest is greater than the growth rate

10. Although the Social Security system holds assets in an accounting trust fund, it is most accurately described as an unfunded PAYGO system (Feldstein and Liebman 2002).

11. This follows from $d_t = \frac{D_t}{Y_t} = \frac{L_t - A_t}{Y_t} = \frac{L_{t-1}(1+r) + NC_t - B_t}{Y_{t-1}(1+g)} - \frac{A_{t-1}(1+r) + C_t - B_t}{Y_{t-1}(1+g)} = \frac{d_{t-1}(1+r)}{(1+g)} + nc_t - c_t$, where D_t is the level of the implicit pension debt; C_t, B_t, NC_t, and Y_t are the nominal period t levels of the annual pension contribution (from both the government and workers), benefit payment, normal cost, and GDP, respectively; and L_t and A_t are liabilities and assets, respectively, at time t. Setting $d_t = d_{t-1}$ and solving for c_t yields equation (1). Here we have assumed that assets and liabilities are subject to the same interest rate r, an assumption that is relaxed in section V and in some of our projections.

of the economy, $r > g$, contributions have to be sufficient to cover the normal cost and the service costs on the implicit debt in order for the implicit debt to be stable as a share of GDP. A plan that makes this level of required contributions will be stable even if it is less than fully funded. Thus, full funding is not required in order for pensions to be fully able to meet their obligations.

An important element of the pension sustainability equation is the dependence on $r - g$. If the rate of interest and GDP growth are equal, $r = g$, and the annual contribution to the pension fund equals the normal cost—the pension equivalent of a balanced primary budget—then the existing stock of implicit pension debt can be maintained as a share of GDP at no fiscal cost. If $r < g$, then implicit debt can be held constant as a share of the economy with contributions less than the normal cost. Of course, the lower is r, the higher is both the pension debt and the normal cost. On net and holding g constant, the lower is r, the more costly it is to achieve long-run pension sustainability.

I.B. Optimal Funding and Intergenerational Equity

In sharp contrast to the emphasis on full funding in most policy discussions of pensions, the theoretical literature on optimal pension funding is decidedly mixed in its conclusions. For example, tax smoothing considerations may dictate a wide range of optimal funding levels, including levels substantially below full funding, depending on economic conditions (D'Arcy, Dulebohn, and Oh 1999). If most voters are borrowers and government borrowing costs are lower than voters' borrowing costs, then no prefunding is optimal in many instances and can be viewed as the logical benchmark (Bohn 2011).[12] Furthermore, to the extent that state and local government expenditures are investments (e.g., schooling) rather than consumption, borrowing is appropriate as the benefits from that spending accrue in the future (Sheiner 2021). Other papers focus on the costs of not prefunding: asymmetric information between government employees and other voters over the cost of pensions may allow government workers to accrue rents in the absence of prefunding (Bagchi 2019; Glaeser and Ponzetto 2014); unfunded pensions may lower the capital stock (Feldstein 1974). Finally,

12. Bohn (2011) observes that most US taxpayers are net borrowers and argues that if borrowing entails intermediation costs—if there is a wedge between financial asset returns and the cost of borrowing—then zero funding is optimal for taxpayers who hold debt. Instead of paying taxes to prefund pension obligations, borrowers are better off paying down their debt because doing so yields a higher return than the market return earned on assets held in a pension fund.

Lucas (2017) provides a thorough discussion of both the uncertainty surrounding optimal funding levels for state and local pensions, as well as arguments for and against full funding.[13]

I.C. Related Literature

This paper is related to a number of recent efforts to examine the fiscal health of public pension plans on an ongoing, forward-looking basis—an area that represents a gap in the large literature on public pensions (Novy-Marx and Rauh 2014a). These papers examine the ongoing flow of future pension obligations, account for the entry of new workers, and explore different paths for asset returns. Novy-Marx and Rauh (2014b) estimate the increase in contributions that would be required for plans to achieve full prefunding under risk-free discount rates over a thirty-year horizon. Although the methodology employed in their paper is broadly similar to that used in portions of this paper, the research questions they asked differ markedly. Based on the logic articulated above, we examine the stress associated with stabilizing a plan's current pension debt. The different questions yield different answers. Novy-Marx and Rauh (2014b) conclude that the cost of transitioning to full prefunding over thirty years is extremely high in most cases and imply a fiscal burden that would very reasonably be called a crisis. In contrast, our analysis concludes that some plans are currently sustainable over the long run and many others can be rendered sustainable at moderate fiscal cost.

Boyd, Chen, and Yin (2019), Boyd and Yin (2016b, 2017), Yin and Boyd (2019), and Shoag and Farrell (2017) allow for stochastic asset returns. They examine the effect of different funding policies, all of which aim to transition to full prefunding, on the future fiscal position of a single, representative pension plan. All conclude that under stochastic investment returns, a wide range of future funding levels is possible. Munnell, Aubry, and Hurwitz (2013) also simulate the effect of stochastic investment returns on future funding status and reach similar conclusions. Mennis, Banta, and Draine (2018) provide stress tests for pension systems in ten states under various asset return assumptions, including stochastic asset returns; their work is related to our calculations for asset exhaustion dates. Boyd and

13. As emphasized in Lucas (2017, 20), in a frictionless, perfect market the degree of funding does not matter to taxpayers or beneficiaries; ultimately, only the size and incidence of the obligations matter. However, in the presence of market imperfections "funding decisions have real consequences. Considerations affecting the best choice of funding rules include intergenerational equity, expectations about future economic growth, optimal tax policy, transparency, fiscal constraints and political incentives."

Yin (2016a) consider the influence of demographic characteristics on the funding levels of five pension plans; this work is related to our examination of the effect of population aging on pension finances. Although he does not examine pensions on an ongoing, forward-looking basis, Rauh (2017) calculates the contribution needed in the current fiscal year to prevent the unfunded pension liability from rising in the next fiscal year. This exercise has some relation to our calculations of the increase in contributions that would stabilize implicit pension debt at its current level. Finally, Costrell and McGee (2020) discuss this paper's pension debt stabilization framework with a focus on asset return risk and provide a stochastic analysis of debt stabilization for the California State Teachers' Retirement System (CalSTRS) plan.

II. Data and Sample Selection

We obtain data from multiple sources on pension plans as of fiscal year 2017. A principal source is the Public Plans Database (PPD) maintained by the Center for Retirement Research at Boston College.[14] The PPD contains plan-level data accounting for 95 percent of state and local pension plan membership and assets in the United States.

The next two major sources of data are the actuarial valuations (AVs) and Comprehensive Annual Financial Reports (CAFRs) for the individual state and local plans in our sample for fiscal year 2017. These documents provide the necessary information required to construct reasonable projections of the plan's liabilities and benefit cash flows. Specifically, for each state we collect the following matrices and distributions: (1) the age and service distribution of currently employed members (actives), (2) average salaries by age and service for the currently employed members, (3) the age distribution of current beneficiaries, (4) the distribution of average benefits for current beneficiaries by age, (5) mortality assumptions by status (active employee or beneficiary), (6) termination rates by age and service, and (7) retirement rates by age and service and plan tier.[15]

The AVs and CAFRs provide further critical information relating to plan provisions and actuarial assumptions not available in the PPD: the plan benefit factor, normal retirement age, early retirement age, service requirement,

14. "Public Plans Data," Center for Retirement Research at Boston College, Center for State and Local Government Excellence, and National Association of State Retirement Administrators, https://publicplansdata.org/.

15. Termination rates include all non-mortality and disability-related causes of employment termination.

vesting requirement, salary averaging method, penalty factor for early retirement (percentage reduction per year early), plan marriage and spousal benefit assumptions, gender ratio of the active employee population, and cost-of-living adjustment assumptions (COLAs).[16] We collect this set of information for each plan tier, where each tier has different parameters for employees, typically depending on date of hire. For instance, tiers within a plan might offer different benefit factors and have different normal retirement dates. (Introducing a new tier is a principal mechanism through which plans have enacted reforms in recent years.) Finally, mortality assumptions are from the Society of Actuaries (SOA).[17]

We estimate the future annual benefit cash flows for a representative set of forty state and local government pension plans. Our sample includes the largest twenty public pension plans in terms of liabilities in the PPD. Our remaining twenty plans are chosen such that our sample matches the national PPD sample in terms of the first and second moments of five plan characteristics measured as of the 2017 fiscal year: the funding ratio (ratio of assets to accrued liabilities calculated using the plan's chosen discount rate), ratio of the unfunded liabilities to current payroll, ratio of current employer pension contribution to payroll, ratio of active plan participants to current beneficiaries, and predicted population growth. The first two characteristics capture how well funded the plan is, the third captures the current budgetary burden of the pension plan, and the final two capture demographic aspects of the plan.

As displayed in table 1, our sample of plans matches the national PPD sample of plans well, both in terms of means and standard deviation; this holds for both unweighted and weighted samples.[18] Our targeting of the

16. Annual pension benefits are typically equal to the years of service multiplied by final average salary times the benefit factor. Thus, the benefit factor is the percent of final salary to which a pension beneficiary is entitled for each year of service. Typically, the average salary from the highest three to five years is used to determine the final salary.

17. Specifically, we use the SOA's RP-2014 Mortality Tables. We also use the accompanying mortality improvement assumptions (Scale MP-2016) to reflect improving mortality rates over our projection.

18. Our sample is selected as follows: we randomly select twenty plans from the PPD and add these to the largest twenty plans from the PPD in terms of stated liabilities to obtain a sample of forty plans. We then calculate the sum of squared deviations between the sample and the PPD universe for the ten targeted moments—that is, the mean and standard deviation of the five plan characteristics. We iterate five thousand times and take the sample with the lowest sum of squared deviations. For this procedure, the five plan characteristics are first transformed to z-scores with mean equal to zero and a standard deviation of one. Thus, the five plan characteristics can be viewed as having equal weight in the sample selection process.

Table 1. Estimation Sample of State and Local Pension Plans

	Unweighted		*Weighted*	
	Estimation sample	*PPD national sample*	*Estimation sample*	*PPD national sample*
Assets/liabilities	0.71	0.72	0.71	0.71
	(0.16)	(0.16)	(0.17)	(0.17)
Unfunded liabilities/payroll	2.38	2.36	2.04	2.00
	(1.69)	(1.81)	(1.59)	(1.60)
Total pension contributions/payroll	0.29	0.30	0.24	0.25
	(0.13)	(0.16)	(0.10)	(0.12)
Active members/retired members	1.31	1.27	1.35	1.35
	(0.37)	(0.41)	(0.34)	(0.34)
Projected active member growth	0.28	0.34	0.41	0.41
	(0.54)	(0.55)	(0.59)	(0.55)
Observations	40	179	40	179

Sources: Public Plans Data, https://publicplansdata.org/public-plans-database/; authors' calculations.

Note: Means are displayed, with standard deviations in parentheses. In the two right columns the samples are weighted by the denominator of the plan characteristics for the first four characteristics (e.g., assets/liabilities is weighted by liabilities). Projected percentage for active member growth is weighted by the number of active members.

second moment of the plan characteristics yields a sample that includes plans with a relatively strong prefunding position, as well as those with a relatively weak prefunding position. For instance, our sample includes the Oklahoma Police Pension and Retirement System and the New York State Teachers' Retirement System, both of which are essentially fully prefunded (using the plans' chosen actuarial assumptions). It also includes the State Retirement Systems of Illinois and the New Jersey Teachers' Pension and Annuity Fund, which have a ratio of assets to liabilities of roughly 40 percent using the plans' assumptions. Our sample also includes many typical plans, such as the Teachers Retirement System of Georgia and the San Diego County Employees Retirement Association, both of which have a funding ratio around 75 percent. Online appendix table B1 provides a complete list of plans in our sample; online appendix table B2 summarizes the inputs for each plan; and finally, as shown in online appendix figures B1 and B2, our sample also matches the national PPD dynamically in terms of mean plan characteristics.

Our use of a sample of plans, as opposed to the universe of plans, reflects the large number of state and local pension plans in the United States—over six thousand according to census data—and the extremely labor-intensive nature of reverse engineering the cash flows. Relative to Novy-Marx and Rauh (2011) we conduct a more detailed, plan-specific reverse engineering

of the cash flows; in particular, we use plan-specific distributions, actuarial assumptions, and benefit information (e.g., normal retirement age). Our modeling of plan tiers, which allows us to assess the effects of recent pension reforms, is a further distinguishing factor. Moreover, we have invested considerable effort into accurately modeling each of our forty plans on a case-by-case basis; for example, in a number of cases we have consulted with the plan administrators or the actuarial firm responsible for the annual actuarial reports in order to resolve uncertainty. Novy-Marx and Rauh (2011), on the other hand, have a significantly larger sample of 116 plans.[19] The different approaches reflect the different aims of the respective papers: ours to estimate the future benefit streams as accurately as possible, in particular their time-varying trajectory; theirs to get the overall liability of pension obligations for the entire state government sector.

III. Methodology

Our methodology for estimating pension fiscal sustainability can be divided into three stages: First, we reverse engineer the future flow of benefit payments to current workers and beneficiaries using plan-specific data and assumptions and the methodology developed by Novy-Marx and Rauh (2011). We use calibration factors to ensure that these cash flows replicate the stated liabilities in the relevant actuarial reports. We then reestimate these cash flows using our own economic assumptions uniformly across plans. Second, we project future plan membership growth and then use our economic assumptions and plan-specific benefit parameters to estimate benefits for future workers using the same methodology as used for current workers. Finally, we pair the benefit cash flow projections with information on plan assets and our own assumption for discount rates and asset returns to assess the fiscal stability of each plan.

III.A. Estimating Cash Flows for Current Workers and Beneficiaries

To construct the cash flows for current beneficiaries and workers, we first collect the data, inputs, and actuarial assumptions discussed in section II for each plan. For current beneficiaries, we then use the mortality tables to age the initial distribution of the beneficiaries each year and use the information on current beneficiaries' pension benefits by age to calculate annual benefit payments. For current workers, we age the workforce each year (incrementing years of service and age) and use the probabilities of

19. Subsequent works by these authors have even larger sample sizes; for example, Novy-Marx and Rauh (2014a) has a sample of 193 plans.

retirement, disability, death, and quits or termination by age and years of service to create a matrix of new beneficiaries by year. We then use the information on pension eligibility, benefit formulas, and economic assumptions to calculate the pension obligations for future beneficiaries by year. These benefit formulas vary by plan tier to capture the effects of reforms implemented between cohorts of active workers.

In order to ensure our projections are as accurate as possible we calibrate our projected cash flows such that they produce each plan's stated actuarial liabilities (AL)—the present discounted value of projected future pension benefits earned to date—as reported in their actuarial valuations. We calibrate separately for current workers, current inactives (individuals who are no longer employees but remain eligible for pensions in the future), and current retirees.

Although these procedures are conceptually quite straightforward, the actual implementation is substantially more complex. Indeed, the challenging and time-consuming nature of the reverse engineering methodology has almost certainly inhibited research on state and local pensions. Our specific procedures for calculating liabilities, which generally follow Winkelvoss (1993), and our calibration methods are presented in detail in online appendix A. Our uncalibrated estimates were on average quite accurate, so the calibration process does not have a large effect on our aggregate analysis (see online appendix table B3).

Finally, we reestimate the future benefit flows using our own economic assumptions. We assume the same rates of change of overall nominal wage growth (3.4 percent) and CPI inflation (2.2 percent) for every plan.[20]

III.B. Methodology for Estimating Benefits for New Hires

In order to study the fiscal stability of each plan we also need to estimate benefit cash flows associated with hires made after 2017. New hires in year t (nh_t) are set equal to the previous year's head count (ee_{t-1}) multiplied by the sum of the projected growth rate in the government's workforce (n_t) and the proportion of withdrawals and retirements from the workforce in the previous year (q_{t-1}).

$$nh_t = ee_{t-1}\left(n_t + q_{t-1}\right) \tag{2}$$

20. These assumptions are consistent with productivity growth of 1.4 percent and a GDP deflator of 2 percent. Our assumption of 2.2 percent annual inflation, as measured in the CPI, is consistent with the Federal Open Market Committee's (FOMC) 2 percent inflation target which pertains to the Personal Consumption Expenditures (PCE) price index. CPI inflation tends to systematically run above consumer inflations as measured by the PCE price index (Haubrich and Millington 2014).

Projected workforce growth (n) is assumed to equal the growth in the working-age population of the state or locality such that the ratio of the government workforce to the working-age population remains constant. We further assume that the age distribution and relative salaries of new hires match the distribution of current employees with fewer than five years of service. Each group of new hires then produces a new stream of benefits starting at each future year, with the value of those future benefits calculated in exactly the same way as they were for the current active workers but adjusting for changes to plan provisions (reforms) instituted for new hires.

To project the growth of the working-age population in each state, we employ a variant of the methodology used by the Demographics Research Group at the University of Virginia's Weldon Cooper Center for Public Service. This methodology projects population by age bins using trends in fertility and in and out migration by state. Our implementation assumes that state population growth eventually converges to the national average—we don't allow states to lose population over the long run, but we do over the medium run in keeping with past trends. In order to calculate state labor force growth rates, we multiply the working-age population in each state by age group by the projected national labor force participation rates by age from the CBO's 2019 long-term budget projection.[21] See online appendix D for details. Finally, we calculate total cash flow streams for a given plan by summing the annual flows for beneficiaries, inactive, actives, and new hires.

III.C. Methodology for Determining Current Assets

We use data on plan rules and demographics from fiscal year 2017 and project benefit flows forward from that point. However, there have been significant changes in asset values and interest rates since 2017. Accordingly, we update each plan's asset valuation to the end of fiscal year 2021 and also base our asset return and discount rate assumptions on financial market data from early in the calendar year 2021.

We update the market value of plan assets using the plan's most recent financial report (fiscal year 2019 for most plans and fiscal year 2020 for some plans). Then, to calculate rates of return since the last observed asset valuation to the present (February 12, 2021), we use the asset allocations

21. For the county- or municipal-level plans we adjust the state projection by the ratio of the growth rate of the local population to the state population over the period 2010–2018. We then phase out this adjustment linearly over time such that by 2050 the locality is growing at the same rate as the state population. See CBO (2019, table 10).

in the financial reports matched to market rates of return on appropriate indexes (see online appendix E). Finally, we use the assumed general asset rate of return—see section III.D—to grow assets from the present to the end of the 2021 fiscal year. On average, we calculate that plan assets will have increased 23 percent since the end of fiscal year 2017.

III.D. Asset Returns and Discount Rates

In order to calculate asset exhaustion dates it is necessary to assume a rate of return on plan assets. The rates of return assumed by plans is typically the expected value of returns on the plan's portfolio of assets. In practice, asset returns in any given year will likely be higher or lower than the long-term average. We primarily present our results using three deterministic asset return assumptions, including the expected rate of return. To address uncertainty in market performance, we also use rates which are risk adjusted downward; these include a risk-adjusted return based on recent market prices.

The question of whether the cash flows in government budget projections *should* be risk adjusted is difficult and contentious.[22] As noted by Kamin (2013), it is not standard practice to risk adjust budget projections. For example, the CBO projects expected revenues and expenditures over time, even though those cash flows are risky. (For example, consider the fact that taxes on capital income are a form of asset ownership.) Using market-based prices to risk adjust the cash flows would lead to much larger deficits and debt. But the goal of stabilizing the federal debt using CBO's non-risk-adjusted projections is widely accepted, even though it leaves future generations with more risk.[23] Furthermore, assuming lower-than-expected rates of return means that, on average, projections will be biased. That is, if the expected return on pension assets is 5 percent, but we assume a return of 2 percent, then we will, on average, underpredict investment returns and overpredict asset exhaustion.

On the other hand, risk adjustment prevents plans from appearing healthier simply because they invest in riskier assets. That is, to the extent expected

22. This issue is related to, but not equivalent to, the contentious issue of the correct discount rate to calculate pension liabilities.

23. On the other hand, for credit and loan guarantee programs like student loans, which under the Federal Credit Reform Act of 1990 (FCRA) are scored on a net present value rather than cash flow basis, the CBO does advocate risk adjusting (Lucas and Phaup 2008; Marron 2014). Official estimates of the costs of federal loan programs are not risk adjusted, but the CBO's preferred measure, which they call fair value, is. For a discussion in support of fair value, see Lucas (2012); for a discussion of the pros and cons of risk adjusting, see Sastry and Sheiner (2015).

cash flows increase simply because the assets have become riskier, the plan would see no benefit when scored using a risk-free rate of return. Furthermore, if the risk-adjustment factor reflects the trade-off taxpayers (current and future) would make between a risky stream and a certain one, then future taxpayers should be indifferent between the cash flows pension plans receive on a risky asset and the cash flows they would receive if the fund invested in safe assets like Treasuries.

In addition, even if one chooses to risk adjust, it is unclear whether the market rate of return on safe assets is the appropriate risk-free rate for government-sponsored pension plans. First, the wedge between the return on Treasuries and riskier investments doesn't only represent risk—it also includes a convenience yield, reflecting the liquidity value of Treasuries and their usefulness as collateral, among other things. Krishnamurthy and Vissing-Jorgensen (2012) estimate that the convenience yield averaged 73 basis points between 1926 and 2008.[24] Second, as noted by Falkenheim (2021), to the extent the risk premium reflects business cycle risk, the government can lower that risk by spreading it across future generations. If the amount of intergenerational risk spreading is less than optimal, then the market risk premium is overstating the cost of risk when borne by the government (and hence future taxpayers). Relatedly, as noted by Sastry and Sheiner (2015), there are benefits to government holding assets that perform well in good times and poorly in bad. If private investors react to temporarily low returns by reducing consumption, but government does not, then government ownership of risky assets may lessen the severity of economic downturns. In that case, taxpayers likely would not be indifferent between pension plans holding Treasuries and risky assets.

The issue of risk adjustment in government accounting is an important one, but settling it is well beyond the scope of this paper. Accordingly, we present our estimates using a variety of real long-run rates of return on the pension assets: a risk-free real rate of 0.5 percent, a real rate of 2.5 percent, and a real rate of 4.5 percent.

The 0.5 percent real rate of return is roughly equal to the longer-run risk-free rate (putting aside the issues discussed immediately above) in recent years. Thus, it represents the rate of return that pension plans can achieve with certainty today, based on financial market prices in recent years—that is, it is the risk-adjusted or risk-neutral rate of return. We obtain the risk-free rate from the yield on the zero coupon twenty-year Treasury Inflation

24. Our market-based measure of the risk-free rate, which is based on Treasury Inflation Protected Securities (TIPS) yields, may be less affected by this because TIPS are less liquid than other Treasuries.

Projected Securities (TIPS).[25] It is worth noting that the current rate is unusually low by historic standards: indeed, CBO long-run economic projections have the real rate on ten-year Treasuries rising over time, reaching 0.9 percent by 2030 and 2.7 percent by 2051 (CBO 2021). Thus, we view the 0.5 percent rate as very conservative—it is quite plausible that plans will be in better shape over time simply because this rate rises.

The 4.5 percent return reflects the expected real rate of return on a pension portfolio comprised of 20 percent risk-free assets and 80 percent equities. The risk-free assets earn the 0.5 percent risk-free rate and the equities earn this rate of return plus an equity (or risk) premium of 5 percent.[26] This 4.5 percent expected real rate of return is equal to about what the plans are assuming on average and slightly less than what they have received on their assets, on average, over the past fifteen years. The 2.5 percent rate of return is equivalent to a mixed portfolio containing 60 percent risk-free assets and 40 percent equities.[27] An alternative interpretation of these asset return assumptions is to view them as capturing realized asset returns in different future states of the world.

In all cases we discount plan liabilities using the 0.5 percent real risk-free rate. This assumption implicitly defines the liability as the amount one would have to pay a private investor to take on the risk. It incorporates the assumption that pension obligations will be paid out in full in nearly all future states of the world and that the value of the payouts (which depends on wages) is uncorrelated with the state of the economy. Neither of these conditions is likely to be strictly true; thus, we view this as a conservative assumption.[28] In any case, as we explain below, our results are not very

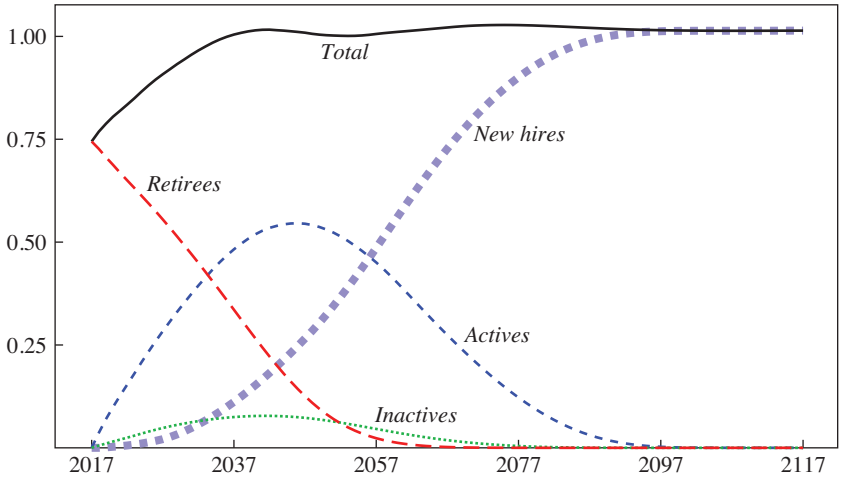
25. The yield on the zero coupon twenty-year TIPS averaged roughly 0.5 percent from the start of 2018 through February 12, 2021. (February 12, 2021, is the date used to obtain financial market data with which to adjust pension plan assets to current values.) The TIPS yield is based on the methodology of Gurkaynak, Sack, and Wright (2008) and obtained at https://www.federalreserve.gov/data/yield-curve-tables/feds200805.csv.

26. We view the 5 percent equity premium assumption as relatively conservative. For example, Duarte and Rosa (2015) estimate that the equity premium has exceeded 10 percent in the years following the Great Recession; Mehra and Prescott (2003) estimate an equity premium of around 7 percent for the United States in the twentieth century; and Novy-Marx and Rauh (2011, 2014a) use an equity premium of 6.5 percent for analyzing pension outcomes.

27. Lucas and Zeldes (2009) discuss the *optimal* asset allocation for state and local pensions and demonstrate that a higher-risk, higher-return allocation of assets can be desirable when it causes distortionary tax rates on average to be lower and when it provides a hedge against liability risk. Nonetheless, the authors conclude that these arguments fail to justify the very elevated share of high-risk assets in most state and local pension portfolios.

28. In particular, most pension plans have the legal ability to change the cost-of-living adjustment (COLA) even for existing retirees.

Figure 2. US Aggregate Ratio of Beneficiaries to Active Workers



Source: Authors' calculations.

Note: The solid line displays the ratio of total beneficiaries of state and local government pension plan payments to the state and local government current workforce. The dashed line labeled "Retirees" displays the ratio of beneficiaries who were receiving benefits as of 2017 to current workers. The dashed line labeled "Actives" displays the ratio of beneficiaries who were employed by a state and local government as of 2017 to current workers. The dashed line labeled "Inactives" displays the ratio of beneficiaries who were no longer employed as of 2017 and who were eligible for a pension benefit, but who had not started to receive the benefit as of 2017 to current workers. The dashed line labeled "New hires" displays the ratio of beneficiaries who were hired after 2017 to current workers.

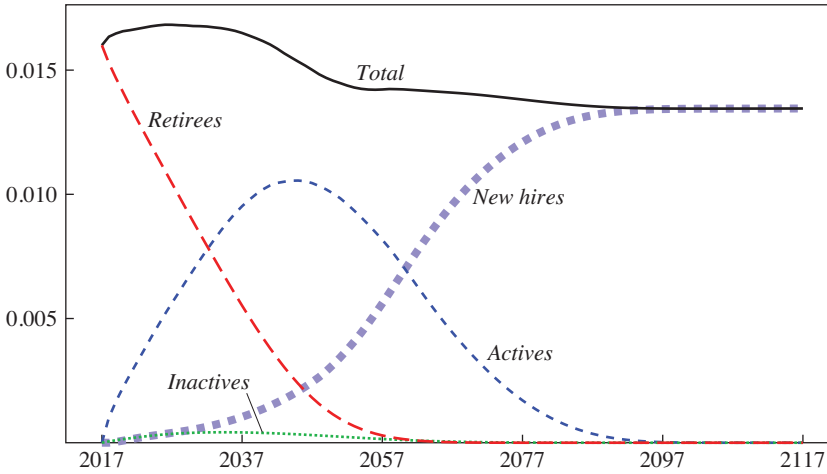
sensitive to the chosen discount rate because we are focusing on stability of the implicit debt rather than its level. That said, exercises that calculate what is required for plans to be fully funded are very sensitive to this assumption.

IV. Results under Current Funding Levels and Benefit Parameters

IV.A. Pension Benefit Payments

Figure 2 shows how the ratio of beneficiaries to active workers evolves over time for our set of plans. The top solid line shows the total, while the dotted lines show the composition. In year 2017—the starting point for our simulation—beneficiaries are just current beneficiaries, but over time, current beneficiaries die, while current workers and current inactive members retire. Meanwhile the workforce is being populated with new workers, and eventually these new hires retire as well.

Figure 3. US Aggregate Ratio of Benefit Payments to GDP



Source: Authors' calculations.

Note: The solid line displays the ratio of total state and local government pension benefit payments to GDP. The dashed line labeled "Retirees" displays the ratio of benefit payments to beneficiaries who were receiving benefits as of 2017 to GDP. The dashed line labeled "Actives" displays the ratio of benefit payments to beneficiaries who were employed by state and local government as of 2017 to GDP. The dashed line labeled "Inactives" displays the ratio benefit payments to beneficiaries who were no longer employed as of 2017 and who were eligible for a pension benefit, but who had not started to receive the benefit as of 2017 to GDP. The dashed line labeled "New hires" displays the ratio of benefit payments to beneficiaries who were hired after 2017 to current workers.

The ratio of beneficiaries to workers in state and local governments is projected to increase about 36 percent from 2017 to 2040 and then roughly stabilize. In comparison, projections by the Social Security actuaries show that, for the United States as a whole, the ratio of Social Security beneficiaries to workers is projected to rise about 39 percent over this time period.[29] We view this similarity as indicating that we have adequately modeled, in aggregate, the future flow of state and local government employees.

Figure 3 shows the annual benefit payments as a share of GDP for the plans in our sample in aggregate, which we refer to as the "US plan" and view as a reasonably good proxy for the state and local pension system in the United States as a whole. In 2017, pension plan benefit payments were

29. These calculations refer to data for Old Age, Survivors, and Disability Insurance (OASDI); "The 2020 OASDI Trustees Report," Social Security, https://www.ssa.gov/OACT/TR/2020/index.html. This is an appropriate comparison because state and local pensions also cover disability as well as retirement.

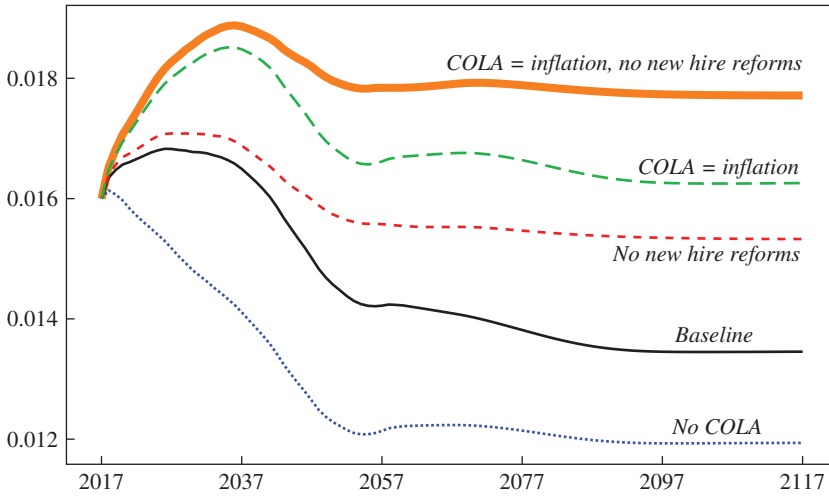
approximately 1.6 percent of GDP. Strikingly, figure 3 indicates benefits are already nearing their peak, rising only about 5 percent over the next ten years before declining to 1.4 percent of GDP by 2070 (about 13 percent lower). This pattern is surprising given the pattern of aging described above. Social Security benefits relative to GDP are projected to rise 21 percent between 2017 and 2040, and then remain roughly constant thereafter.

What explains these surprising results? If the ratio of beneficiaries to workers is increasing, why isn't the ratio of benefits to GDP? First, most pension plans do not fully index their retiree benefits for inflation—the cost-of-living adjustment (COLA) is often well below inflation. Many plans have been lowering or eliminating their COLAs in recent years and this lowers the real value of average benefits over time. Specifically, since 2007, twelve plans in our sample have legislated changes making their COLA less generous or even eliminating it. A further five plans have been able to lower their COLA by reducing or eliminating supplemental or ad hoc COLAs.[30] Second, pension plans have gradually been making changes over time to lower benefits and raise retirement ages for new hires (Aubry and Crawford 2017). These adjustments also reduce average pension benefits over time. The reduced growth in average benefits due to the new hire reforms and changes to COLAs offsets a large share of the effects of the 36 percent growth in the ratio of beneficiaries to workers shown above.

Figure 4 again presents our baseline estimate for benefits payments as a share of GDP, as well as several counterfactual exercises which explore the effect of policy changes. The lowest line displays the aggregate cash flows assuming that plans turned off their COLAs entirely, which governments generally (but not universally) can do without violating state constitutions. The result of eliminating the COLAs would be a drop in the ratio of benefits to GDP, such that they would eventually settle an additional 15 percent below where we project them when the current COLAs are maintained, and about 26 percent below their level in 2017. In contrast, consider the line displaying the results of setting all COLAs to equal inflation. Benefit flows rise substantially as a share of GDP over the next two decades and eventually settle at a much higher level—indeed, the rise is about 16 percent, much closer to the 21 percent projected rise in Social Security benefits described above. Clearly, COLAs have a significant impact on benefit flows as a share of the economy. The middle line displays the trajectory of

30. Fitzpatrick and Goda (2020) note that, because new worker reforms take time to yield budgetary savings, many state and local pension plans have turned to COLA adjustments to address funding concerns. They also document that most COLAs in recent years have been downward adjustments.

Figure 4. US Aggregate Ratio of Benefit Payments to GDP under Different Scenarios



Source: Authors' calculations.

Note: The line labeled "Baseline" displays the ratio of total state and local government pension benefit payments to GDP. The line labeled "No new hire reforms" displays the ratio of total state and local government pension benefit payments to GDP assuming that all pension changes which apply only to new hires—that is, new worker reforms—are canceled. The line labeled "COLA = inflation" displays the ratio of total state and local government pension benefit payments to GDP assuming that all plans set their cost-of-living adjustment (COLA) to equal the rate of inflation. The line labeled "No COLA" displays the ratio of total state and local government pension benefit payments to GDP assuming that all plans set their COLA to equal zero.

benefits to GDP when the reforms for new workers are eliminated and we instead assume that new hires are subject to the same pension rules as current workers. Rather than declining by 13 percent over time, the ratio of benefits to GDP would stabilize at about the same ratio as today.[31]

Finally, the top line displays the path of benefits to GDP when both the new worker reforms are eliminated and COLAs are set equal to inflation. In this scenario, benefits as a share of the economy are projected to rise 17 percent between 2017 and 2040—reasonably close to the 21 percent increase projected for Social Security. Thus, new worker reforms and

31. This analysis assumes that these new worker reforms remain in place going forward. Of course, there is a possibility that some of these reforms may be revoked or altered. For instance, the 2010 "tier II" reform instituted for state-administered plans in Illinois has been widely criticized for creating a very significant disparity in benefit generosity for employees hired before and after 2011. Moreover, it is possible that the reform may eventually run afoul of federal law (Bruno, Kass, and Merriman 2019).

COLAs explain the majority of the more muted rise in state and local pension benefits compared to Social Security.[32] Online appendix table A1 presents benefit payments on a plan-specific basis in 2017 and 2047; for 2047 the "no reforms" and "no reforms and COLA equals inflation" counterfactuals are also displayed.

The fact that pension benefits as a share of payroll are, in aggregate, near their highest level expected over the next few decades is an important finding for understanding the sustainability of state and local finances and the ability of plans to smooth through the next few decades. Notably, as displayed in online appendix figure A4, the flattening out of pension benefit payments as a share of GDP is readily apparent in the historical data.

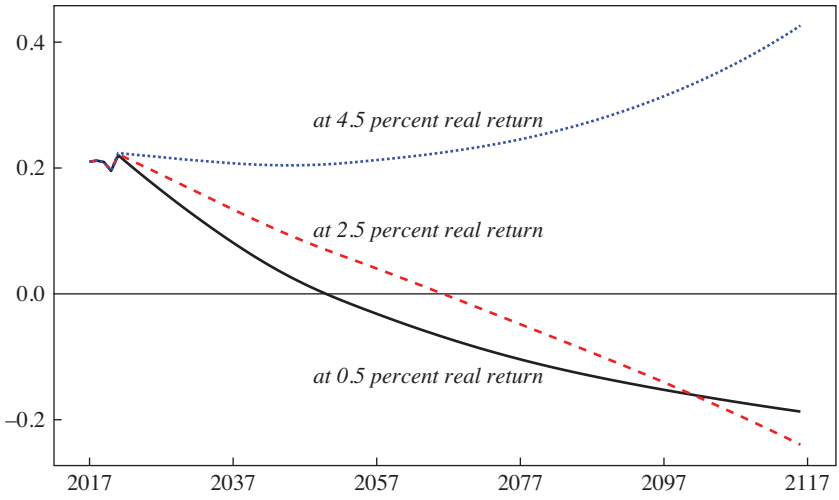
IV.B. Pension Asset Projections

To determine whether plans are fiscally sustainable, we hold the annual contributions of employees and employers (as a share of GDP) fixed at today's level and assume that benefits evolve as described in figure 4.[33] We view this as performing a current policy analysis, akin to the current law baseline used by CBO in its projections for the federal budget. Figure 5 shows the path of pension assets in this current policy analysis under our three asset return assumptions. With the 0.5 percent real rate of return, current contributions are insufficient to keep the plans solvent. Despite the projected decline in benefits relative to GDP, assets relative to GDP begin declining immediately and are exhausted in about thirty years. With a 2.5 percent rate of return, assets are declining, but not as quickly; they are exhausted in about forty-seven years. If, however, the plans earn 4.5 percent on their assets, then plans are sustainable: at current contribution rates, assets rise indefinitely and the plans face no fiscal stress (indeed, one would

32. Other possible explanations for the reduced growth in average benefits, other than changes in COLAs and new worker reforms, include sluggish state and local government wage growth over the past fifteen years, lower average tenure of benefit recipients over time, and a secular transition toward less generous pension plans due to the relative population shift away from the Northeast and Midwest (whose governments tend to have relatively generous pension plans).

33. More precisely, we hold contributions as a share of GDP fixed at its current value for each plan tier. Some plans have employee contribution rates that differ by tier. For these plans, as the composition of the workforce shifts over time away from the tier(s) for longer-tenured employees and toward the tier(s) for shorter-tenured employees, the overall plan contribution rate will shift. This is particularly an issue with plans that have lowered the generosity of their defined-benefit plan while introducing a defined-contribution plan (i.e., plans that have gone hybrid). We account for both the declining contribution rate and the declining generosity of the defined-benefit plans.

Figure 5. US Ratio of Assets to GDP



Source: Authors' calculations.

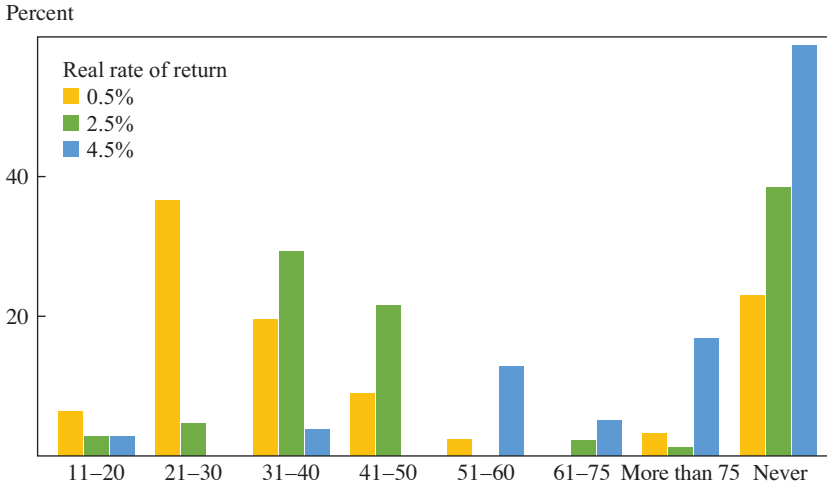
Note: The figure displays pension assets as a share of GDP under varying assumptions about asset returns and assuming that employer contributions as a share of payroll are held fixed at their 2017 value.

argue that current contribution rates are too high, if one could count on a 4.5 percent real rate of return).

Of course, looking at the US pension system as a whole masks a lot of variation across plans. Figure 6 shows what share of liabilities are in plans that exhaust within various time periods. With a 0.5 percent real rate of return, about 6 percent of liabilities are in plans that exhaust within twenty years and 43 percent are in plans that exhaust within the next thirty years; even at this low rate of return, 23 percent of liabilities are in plans that never exhaust. At a 2.5 percent real rate of return, only 7 percent of liabilities are in plans that exhaust within the next thirty years, and 38 percent are in plans that never exhaust. With a 4.5 percent real rate of return, almost 60 percent of liabilities are in plans that never exhaust, whereas the other plans do exhaust, but mostly not for many decades.[34] (Online appendix table A2 reports the exhaustion rates for all the plans under the three rate of return assumptions.)

34. One notable exception is the New Jersey Teacher's Plan, which exhausts in ten years even with a rate of return of 4.5 percent.

Figure 6. Percent of Total Liabilities in Plans That Exhaust Their Assets over Various Time Horizons



Source: Authors' calculations.

Note: The figure displays the share of total pension liabilities held by plans which exhaust their assets over different time horizons assuming that employer contributions as a share of payroll are held fixed at their 2017 value.

The message from these exercises is that, for the majority of plans, there is no imminent crisis in the sense that plans are likely to exhaust their assets within the next two decades. But many plans are not stable and a sizeable share of plans will exhaust their assets within thirty years under the low return scenario. Adjustments may be necessary. The questions are: How large are those adjustments, and how urgent are they?

V. Pension Debt Stabilization

V.A. Pension Debt Stabilization Discussion

Our fiscal sustainability exercises are focused on the following identities concerning the evolution of plan liabilities (L), assets (A), and implicit pension debt (D),

$$L_{t+1} = (1+\delta) L_t - B_{t+1} + NC_{t+1} \tag{3}$$

$$A_{t+1} = (1+r) A_t - B_{t+1} + C_{t+1} \tag{4}$$

where δ is the discount rate used to value the plan liabilities; r is the expected return on assets; B_t is the benefit paid out at time t; NC_t is the total normal cost (the present value of liabilities accrued in a year) in year t; C_t is the total contribution. (The difference between δ and r is discussed below.)

Dividing equation (5) and equation (6) by time $t + 1$ GDP (Y_{t+1}), subtracting $\frac{L_t}{Y_t}$ and $\frac{A_t}{Y_t}$, respectively, and rearranging yields the changes in liabilities and assets as shares of GDP from t to $t + 1$:

$$\Delta l_{t+1} = \frac{(\delta - g)\, l_t}{1+g} - b_{t+1} + nc_{t+1} \tag{5}$$

$$\Delta a_{t+1} = \frac{(r - g)\, a_t}{1+g} - b_{t+1} + c_{t+1} \tag{6}$$

where lower case denotes variables as a share of GDP and g denotes GDP growth.

Debt stability requires $\Delta d_{t+1} = \Delta l_{t+1} - \Delta a_{t+1} = 0$. In steady state, liabilities are constant as a share of GDP, $\Delta l_{t+1} = \Delta l = 0$. Thus, in steady state, $\Delta a = 0$. Setting $\Delta a_{t+1} = 0$ in equation (6) yields the steady-state contribution to stabilize debt at any given asset level and steady-state benefit outflow:

$$c = b - \frac{(r - g)}{(1+g)} a \tag{7}$$

When assets are zero, as in a pure PAYGO system, contributions just have to cover benefits. When $r > g$, a plan with assets can have contributions lower than benefits in steady state, because some of the asset income can be used to pay for benefits (while some must be reinvested in order for assets to rise with GDP). When $r < g$, stabilizing debt and assets to GDP actually becomes more costly the larger the assets.

Note that equation (7) includes only the return on assets and not the discount rate. When liabilities as a share of GDP are constant, as they are in steady state, stabilizing debt implies stabilizing assets, and the trajectory of assets is wholly independent of the rate used to discount liabilities. Thus, the required contribution to stabilize pension debt is independent of the discount rate. That said, liabilities are not constant in all plans over the first few years of our projections, because of demographic changes and changes in plan rules that take time to work their way through to benefits. Thus,

the discount rate assumption we use is not entirely neutral, but the effects of changing that assumption are not economically important.[35]

We use these identities in combination with our projections of benefits cash flows and payroll to assess the fiscal stability of each plan. If pension debt as a share of the economy is declining or stable, then the plan can be viewed as fiscally sustainable; assets will never exhaust, and the plan will be able to pay benefits indefinitely. On the other hand, if debt as a share of GDP rises indefinitely, then the plan is not fiscally sustainable.

V.B. Stabilization Exercises

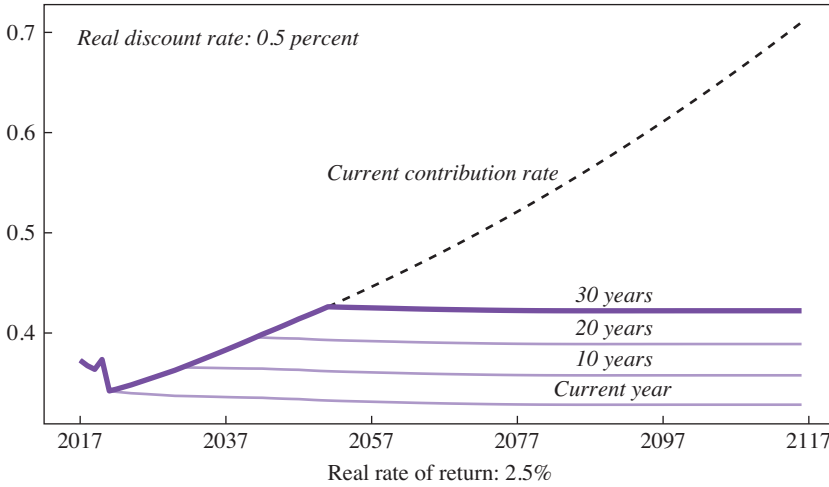
Our analysis here involves estimating the changes in pension contributions which would stabilize pension debt as a share of the economy. We perform two stabilization exercises: First, for long-run stabilization, we ask what onetime and permanent changes in the contribution rate would make implicit pension plan debt eventually stabilize as a share of GDP (without specifying what that share is). Sheiner (2018) does this exercise for the federal debt. Second, for medium-run stabilization, we ask what onetime and permanent changes in contribution would be required in order for the implicit debt as a share of GDP to equal today's ratio in thirty years' time. CBO (2020) does this type of exercise for the federal debt.

STABILIZATION EXERCISE 1: STABILIZE IMPLICIT DEBT AS A SHARE OF GDP IN THE LONG RUN Our first stabilization exercise assumes that a government's pension plan is stable so long as the unfunded liabilities relative to GDP are constant at some point in the future, regardless of the value of this stable ratio. This exercise spreads the fiscal costs of future pension obligations and existing pension debt equally across generations as a share of income.

We first calculate the onetime, but permanent, change in the pension contribution a plan would have to make in order to achieve stability and then

35. In some of our debt stabilization scenarios, we set $r \neq \delta$. Costrell and McGee (2020) criticize this choice, referring to it as "arbitrage," and note that it is a sharp departure from standard actuarial practice where required contributions are constrained to be at least as high as the cost of newly accrued benefits (normal cost) valued at the actuarial discount rate. However, we present results below in which both r and δ are set equal to a 0.5 percent risk-free rate. More fundamentally, this paper focuses on the contribution required to stabilize pension debt and, as discussed above and also by Costrell and McGee (2020), the choice of discount rate has little effect on the contribution required to stabilize pension debt. Thus, harmonizing our asset return and the discount rate would not materially alter our conclusions. Determining the appropriate rate of asset return r, though, has extremely important implications for the required contribution. See section III.D for a discussion. See also Lucas (2017, 3) for arguments in favor of delinking the choice of discount rate used to value pension liabilities from choices over funding requirements, for example "it may be sensible to partially link funding rules to expected return on assets."

Figure 7. US Implicit Pension Debt under Pension Debt Stabilization (Stabilization Started at Different Time Horizons)



Source: Authors' calculations.

Note: The dashed line displays implicit pension debt—unfunded pension liabilities—as a share of GDP assuming that assets have a real return of 2.5 percent and that employer contributions as a share of GDP are held fixed at their 2017 value. The line labeled "Current year" displays implicit pension debt—unfunded pension liabilities—as a share of GDP assuming that assets have a real return of 2.5 percent and that pension contributions as a share of payroll receive an immediate onetime, permanent change such that pension debt eventually stabilizes in the longer run. The lines labeled "10 years," "20 years," and "30 years" are analogous to the line labeled "Current year" but assume that the adjustment to pension contributions occurs in ten years, twenty years, and thirty years, respectively.

assess how that contribution changes depending on whether the government acts now, acts in ten years, twenty years, or thirty years. Figure 7 shows the evolution of the unfunded liability relative to GDP for the United States as a whole if real asset returns are 2.5 percent under the current policy analysis discussed in section IV. The dashed line shows that without changes in contribution rates, implicit debt to GDP rises at an increasing pace over time: the current situation is unsustainable. The other four lines show the trajectory of the debt to GDP ratio if the governments acts now or later. If they act now, the implicit debt to GDP ratio essentially holds steady at around 33 percent in all periods. Waiting to stabilize does not change the steady-state ratio much. If the governments wait thirty years to act—that is, if they maintain their current contribution rate for thirty years and then act to stabilize—the long-run implicit debt to GDP ratio is 43 percent—about 30 percent higher than it would be if the government acted today. The left panel of table 2 presents the contribution increases,

Table 2. Percentage Point Increase in Contribution Rate Required (Percent of Payroll)

	Stabilize implicit debt to GDP				*Implicit debt gets back to today's level in 30 years*				*Fully funded in 30 years*			
Real rate of return	*Today*	*In 10 years*	*In 20 years*	*In 30 years*	*Today*	*In 10 years*	*In 20 years*	*In 30 years*	*Today*	*In 10 years*	*In 20 years*	*In 30 years*
0.5	12.5	11.1	9.9	8.7	14.7	18.2	21.1	23.4	47.2	51.1	54.1	56.2
2.5	6.9	7.4	8.0	8.6	6.2	8.7	11.5	14.3	30.3	33.2	36.1	38.8
4.5	−1.5	−2.0	−2.6	−3.4	−3.0	−4.0	−5.3	−6.7	14.7	13.9	12.7	11.2

Source: Authors' calculations.

Note: The left panel displays the onetime, permanent percentage point change in contributions as a share of payroll required to stabilize implicit pension debt as a share of GDP for the United States in aggregate. The central panel displays the onetime, permanent percentage point change in contributions as a share of payroll required to return implicit pension debt as a share of GDP to today's level in thirty years for the United States in aggregate. The right panel displays the onetime, permanent percentage point change in contributions as a share of payroll required to achieve full prefunding in thirty years for the United States in aggregate.

as a share of payroll, required to stabilize the debt-to-GDP ratio for all four asset return scenarios. At a 4.5 percent real rate of return, plans are, in aggregate, already stable and can lower contributions. At the 2.5 percent rate of return, plans must increase contributions by 6.9 percent of payroll. Under the constant risk-free 0.5 percent return assumption, contributions must increase by a larger 12.5 percent. (Online appendix table A3 presents plan-specific changes in required contributions to stabilize pension debt in the long run; online appendix table A4 reports the associated long-run, stabilized implicit debt-to-GDP ratios and also presents the peak debt-to-GDP ratio reached on the transition to the long-run value.)

The contribution changes required to stabilize implicit pension debt don't change very much if the government waits to make contribution changes. If the contribution rate stays at its current level and then increases in ten years, the increase has to be equal to 7.4 percent of payroll under 2.5 percent asset returns. Acting sooner rather than later lowers the required increase, but not by much. Even if the plans wait thirty years to act (i.e., go thirty years without any changes in contributions), the required increase only rises to 8.6 percent of payroll. Delaying, though, does result in a somewhat higher level of pension debt in steady state. Under the risk-neutral 0.5 percent asset return assumption, required contributions actually fall if a government delays adjustment. This is a striking result—to simply stabilize the debt, there is nothing gained from increasing contributions now. By waiting to act, a plan can contribute much less now, and somewhat less in the future—ultimately stabilizing at a higher debt level at lower future cost. This result follows from the fact that when interest rates are less than the growth rate of the economy, government debt has no fiscal costs (Blanchard 2019). Equivalently, when $r < g$, assets are costly because they constantly shrink as a share of the economy; thus, running down assets and then beginning the stabilization process allows stabilization with a lower contribution rate—see equation (7).[36]

This comparison highlights an interesting conundrum—when asset returns are higher, plans are in better shape and need to contribute less to stabilize their debt. When asset returns are lower, plans are in worse shape as stabilizing pension debt is more expensive, but for both the lowest and even

36. One of our discussants notes that in some theoretical models, r must exceed g. However, $r < g$ is possible in long-run equilibrium in theoretical models with sufficient risk, particularly those that incorporate significant idiosyncratic uninsurable risk, such as heterogeneous agent New Keynesian (HANK) models. Moreover, as noted by Blanchard (2019, 1197), "the current US situation, in which safe interest rates are expected to remain below growth rates for a long time, is more the historical norm than the exception."

middle rates of return there is little benefit to having assets, and so plans are worse off or barely better off by increasing contributions. Of course, assets provide insurance against uncertainty and may allow for smoothing pension contributions over time, and so plans may want to contribute even if there is little benefit when asset returns are certain, as we assume here.

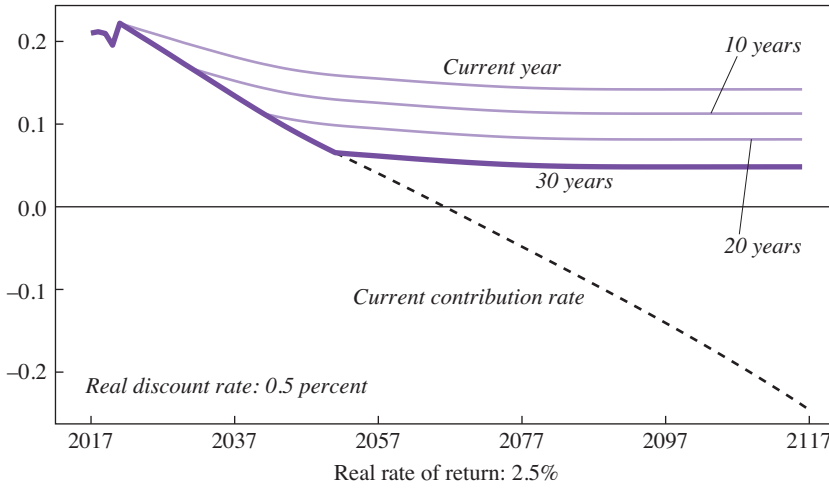
How fiscally onerous would these increased contributions be? To put these contribution changes into context, aggregate pension contributions increased by 10 percent of payroll between 2009 and 2019 and equaled 27 percent of payroll in 2019.[37] Accordingly, if governments act now, a further upward adjustment equal to the adjustment made over the last decade would be more than sufficient to stabilize their pension debt under the 2.5 percent return assumption. Under the 0.5 percent rate of return, plans would have to do more—raising their contribution by 25 percent more than the increase over the past decade. (But, at this low rate of return, plans would be better off not acting in the near term if all they cared about was eventually stabilizing debt.) Overall, we view the contribution changes needed to obtain pension debt stability as achievable, although they would certainly entail some fiscal strain, particularly under the 0.5 percent return scenario.

However, plans could run out of assets along the way, which might be a constraint, both economically—if ratings agencies react by raising borrowing costs—and politically. Figure 8 shows plan assets relative to GDP for each of the 2.5 percent asset return scenarios and illustrates that the long-run stabilization exercise involves plans drawing down assets in order to smooth through the period of peak cash flow demand over the next two decades. Although assets decline in these debt stabilization scenarios, they never approach zero in aggregate. That said, some individual plans do exhaust their assets and stabilize at negative asset values in the various scenarios; and assets fall into negative territory in aggregate in the 0.5 percent for the scenario in which they wait thirty years to stabilize (see online appendix table A5).[38] The simulations yielding negative assets effectively assume that these governments issue marketable debt—akin to the pension obligation bonds which have been issued by some governments in the past—to fund benefits once their assets have been exhausted and thereafter make service payments on the marketable debt at an interest rate equal to the assumed asset return.

37. Based on full PPD sample, updated through fiscal year 2019.

38. For instance, a number of plans in our sample that are poorly funded now and have responded by cutting COLAs or future benefits or both—such as the Illinois State government plans and the New Jersey Teacher's Plan—end up with negative assets in the 2.5 percent scenario.

Figure 8. US Pension Assets under Pension Debt Stabilization (Stabilization Started at Different Time Horizons)



Source: Authors' calculations.

Note: The dashed line displays pension assets as a share of GDP assuming that the assets have a real return of 2.5 percent and that pension contributions as a share of GDP are held fixed at their 2017 value. The solid line labeled "Current year" displays pension assets as a share of GDP assuming that the assets have a real return of 2.5 percent and that pension contributions as a share of payroll receive an immediate onetime, permanent change such that pension debt eventually stabilizes in the longer run. The lines labeled "10 years," "20 years," and "30 years" are analogous to the solid line labeled "Current year" but assume that the adjustment to pension contributions occurs in ten years, twenty years, and thirty years, respectively.

STABILIZATION EXERCISE 2: STABILIZE IMPLICIT DEBT AS A SHARE OF GDP IN THE MEDIUM RUN Another way to assess sustainability is to ensure that the implicit debt-to-GDP ratio is no higher in thirty years than it is today. Very long-run projections are inherently uncertain, so choosing a target implicit debt-to-GDP ratio over the medium-term may be a more reasonable policy objective. In addition, the exercise above that stabilized the implicit debt-to-GDP ratio without specifying its level did not account for potential changes in borrowing costs that might arise if the ultimate debt-to-GDP ratio were higher than it is today—for example, due to credit rating downgrades—whereas targeting today's level is less likely to raise that concern. In addition, a government may wish to simply maintain implicit pension debt in relation to GDP—that is, intuitively dig the hole no deeper while spreading the costs of doing so evenly over thirty years. This exercise is consistent with this objective, on net, over a thirty-year horizon.

The middle panel of table 2 reports the onetime, permanent contribution change required for the implicit debt-to-GDP ratio, at the end of thirty

years, to equal its value in 2021 for the United States as a whole. It should be noted that, in this experiment, we always allow the pension plan thirty years to get back to the original debt ratio, so that starting in ten years means getting back to the 2021 debt-to-GDP level by 2061. We view that as a sensible experiment because it doesn't require the plan to make extremely large changes in a short period of time but still requires the plan to eventually return to the same target.

At a 2.5 percent rate of return on assets, plans would need to increase contributions by 6.2 percent of payroll today, 8.7 percent if they began in ten years, and 11.5 percent if they began in twenty years. There is little difference between the contributions required under this exercise and the long-run stabilization exercise (the left panel in table 2) if action is taken today; but the difference becomes somewhat larger if stabilization is delayed. This difference arises because the thirty-year exercise requires any increases in debt that occur after 2021 to be paid down, whereas the long-run exercise only requires additional interest be paid on debt acquired after 2021. Online appendix figures A5 and A6 show for the United States as a whole the trajectory of implicit debt and assets, respectively, under these stabilization exercises.

At an asset return of 0.5 percent, contributions would have to increase about 15 percent to ensure that the debt-to-GDP ratio is the same as today's in thirty years, just a bit above the amount required in the stabilize the implicit debt in the long-run exercise. However, the differences between the costs of delay across the thirty years and long-run exercises are much larger under these low asset returns, because even though the costs to stabilize a given level of debt are lower the higher that debt, the costs to actually pay down debt are quite high since asset returns are so low. Waiting ten years to take action at the 0.5 percent asset return if plans wanted to ensure that the debt ratio returned to this year's level in thirty years would require an increased contribution of 18.2 percent of payroll; waiting twenty years would boost that required contribution to 21.1 percent (but recall that the plan benefits form the lower contributions over the first twenty years).

Online appendix table A6 presents plan-specific changes in contributions to stabilize over thirty years. Online appendix table A7 presents plan-specific funding ratio estimates for the thirty-year stabilization exercise; the table shows that assets do become negative for a few plans.[39] While many plans do have the ability to issue debt through instruments like pension

39. These are plans with sharply declining liabilities: for debt to remain constant, a decline in liabilities has to be offset by a decline in assets. When the decline in liabilities is large relative to starting assets, maintaining debt can require assets going negative.

obligation bonds, some may not. As an alternative way of assessing stability, we calculated the contributions required for plans to have the same funding ratio in thirty years as they do today. For the United States as a whole, required contributions under this exercise, if the adjustment is made now, are between 2 percent and 4 percent of payroll higher than those required to achieve the same debt in thirty years, depending on the rate of return assumption. These results in aggregate, and by plan, are reported in online appendix table A8.

In contrast to our focus on stabilizing implicit pension debt, past work on pension funding has often focused on achieving full prefunding over a fixed period of time. The rightmost panel of table 2 presents estimates of the funding increase required to achieve full prefunding over a thirty-year horizon. These estimates are broadly similar to those presented in Novy-Marx and Rauh (2014b).[40] The increases required to reach full funding are very substantially larger than those required to stabilize debt. Under 2.5 percent asset returns, the contribution boost to reach full funding in thirty years is roughly four and a half times larger than the increase required to stabilize the debt over the long run (30 percent versus 7 percent). The funding increases required to reach full funding under the 0.5 percent and 2.5 percent return assumptions would be hugely challenging, if not infeasible, for state and local governments.

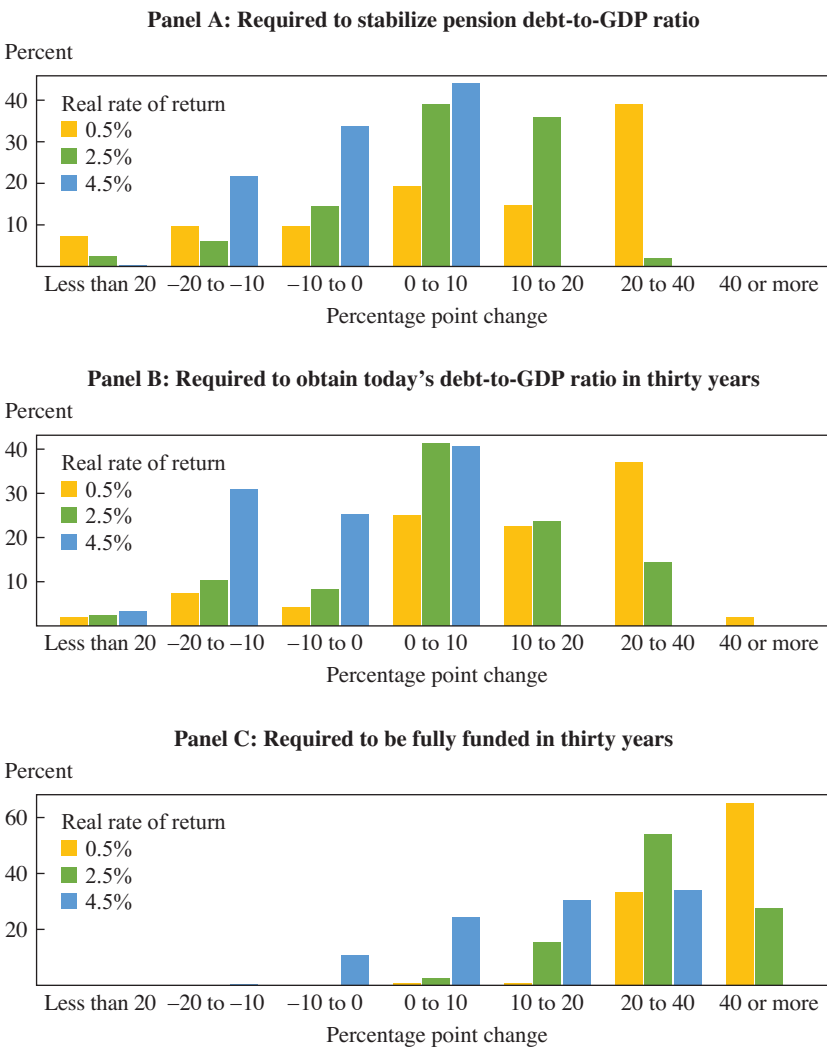
Finally, online appendix A discusses our calculated normal costs; readers interested in comparing the evolution of normal costs over time and in comparing these annual service costs to required contributions under various debt stabilization scenarios should refer to online appendix table A9 and the associated text.

V.C. Variation in Required Contribution Adjustments across Plans

Figure 9 shows the distribution of required adjustments across the asset return assumptions and stabilization exercises. Panel A shows the distribution of required adjustments for plans to stabilize the debt over the long run starting immediately. At a 4.5 percent rate of return, no plan needs to increase funding by more than 10 percent of payroll, and 56 percent of

40. One difference is that our pension liabilities are defined using an accrued liability concept (generally implemented as the entry age normal, or EAN), which includes some benefit obligations associated with future years of service. In contrast, Novy-Marx and Rauh (2014b) mostly use the narrower accumulated benefit obligation concept, which only captures obligations earned to date. Another difference is that our projections include the assumption of mortality improvements over time whereas those of Novy-Marx and Rauh (2014b) do not.

Figure 9. Distribution of Liabilities by Percentage Point Change in Contribution



Source: Authors' calculations.

Note: Figure displays the distribution of liabilities by the percentage point change in contributions (share of payroll) required to stabilize the pension debt-to-GDP ratio, return to today's debt-to-GDP ratio in thirty years, and fully fund in thirty years under different asset return assumptions. The histograms are weighted by liabilities.

liabilities are in plans that could reduce contributions. At the 2.5 percent rate of return, only 2 percent of liabilities are in plans that need to increase funding by more than 20 percent of payroll, and less than 40 percent of liabilities are in plans where the contribution increase is more than 10 percent of payroll. At a 0.5 percent rate of return, however, 39 percent need to increase contributions by more than 20 percent of payroll. Thus, under this rate of return assumption, many plans do have to make significant changes.

Panel B of figure 9 shows the distribution of plans' required contribution changes if they act today for the thirty-year, medium-term stabilization exercise. The distribution is quite similar to the results for the long-term stabilization exercise, although the required contributions are generally a bit larger.

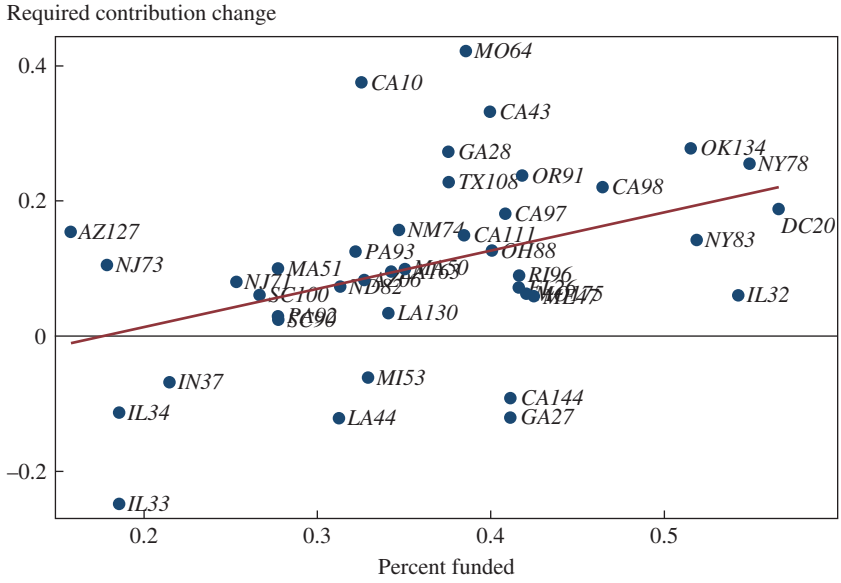
Finally, panel C of figure 9 shows the distribution of required contribution changes for plans to be fully funded by the end of thirty years. At a 4.5 percent rate of return, only 11 percent of liabilities are in plans that can lower contributions, while 34 percent of liabilities are in plans where the required contribution increase is greater than 20 percent of payroll. At the 0.5 percent rate of return, most plans (65 percent) have to increase pension contributions by more than 40 percent of payroll. These comparisons make clear the policy importance of recognizing that pension plans can be stable without being fully funded. An attempt to enact the massive increases in contributions that would be required to move toward full funding at low and moderate asset returns would very likely spark a fiscal crisis. Our analysis, though, demonstrates that increases of this magnitude are unnecessary for plans to become fiscally stable and continue paying benefits.

V.D. Explaining the Variation in Required Contributions to Stabilize the Debt

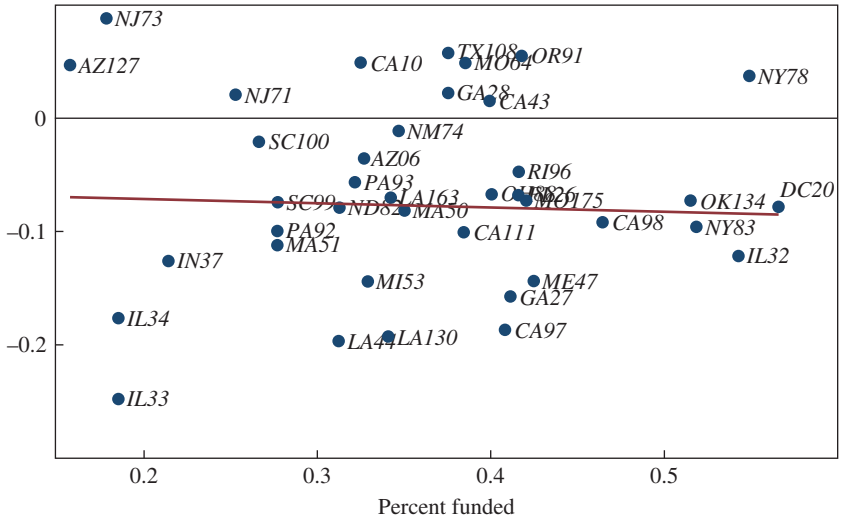
Perhaps unintuitively, it's not the poorly funded plans that have to make the greatest contributions to stabilize. As shown in panel A of figure 10, there is a positive relationship between funding levels and the required contribution change to stabilize under the medium-term stabilization exercise with the 0.5 percent return assumption. At low rates of return, having assets is expensive because the rate of return is not sufficient to keep current assets growing with GDP; rather than being able to use some of the asset returns to fund benefits, plans have to actively contribute to the plan just to prevent assets from eroding. At a 4.5 percent rate of return (panel B), that is no longer the case, but there is little relationship between funding and required contribution. Of course, one reason to expect a relationship between funding level and required changes is that poorly funded plans may be those that

Figure 10. Required Contribution to Stabilize

Panel A: 0.5 percent rate of return, same debt in thirty years

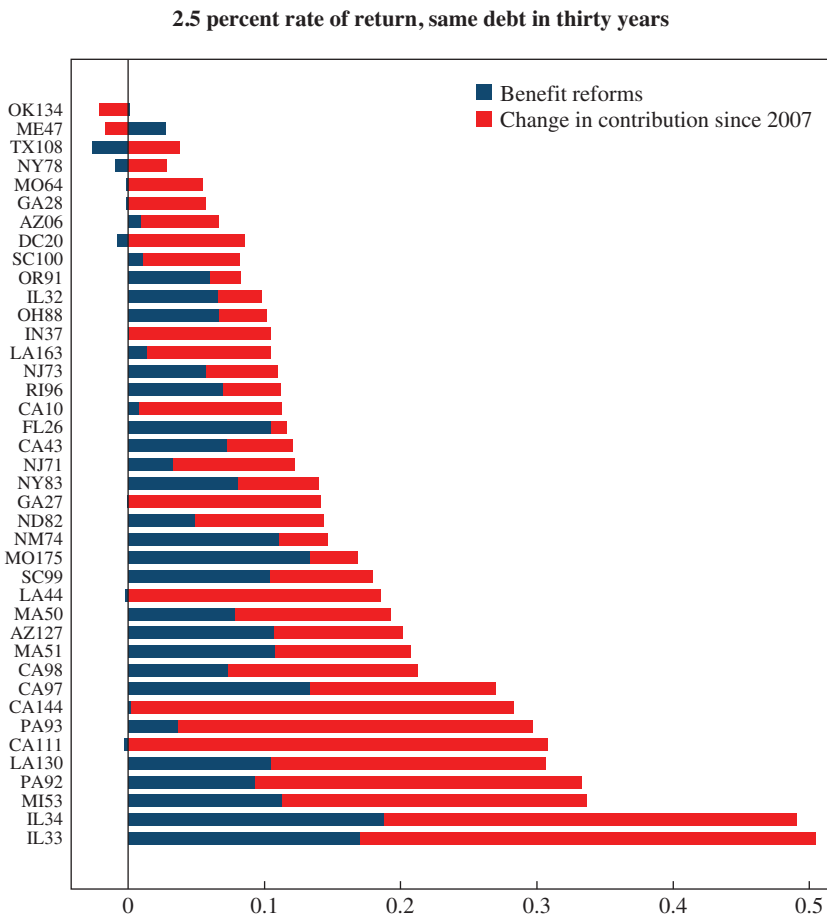


Panel B: 4.5 percent rate of return, same debt in thirty years



Source: Authors' calculations.
Note: The labels are the Public Plans Database's assigned pension plan IDs.

Figure 11. Effects of Changes in Benefits and Contributions on Required Contribution

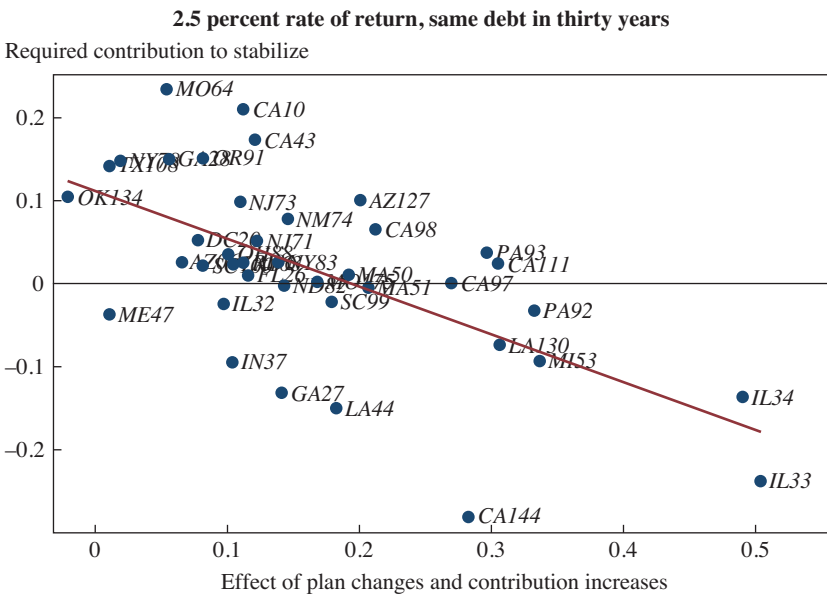


Source: Authors' calculations.
Note: The labels are the Public Plans Database's assigned pension plan IDs.

have been failing to make sufficient payments and have ignored looming imbalances. But that's not the case on average.

Figure 11 shows the effects of recent changes to pension plan contributions and the new-hire reforms discussed above on the contribution change required in the medium-term stabilization exercise. To calculate these, we ran a counterfactual simulation that—starting with today's liabilities and assets—assessed the changes in contribution that would be required to stabilize debt in thirty years if plans reversed the reforms to their benefit and eligibility levels and if their contribution rates reverted to those prevailing

Figure 12. Required Contribution Lower in Plans That Have Made Large Changes



Source: Authors' calculations.

in 2007.[41] We then calculate the difference between those required changes in contributions and the required changes in contributions we calculate under current plan benefits and contributions. The effects of these reforms on the required contribution to stabilize the debt—plotted on the horizontal axis of figure 12—have been substantial. For example, without the reforms and contribution increases made by two of the most poorly funded plans—Illinois teachers and Illinois State employees—required contributions to stabilize the debt in thirty years under the 2.5 percent asset return assumption would equal about an *additional* 50 percent of payroll beyond what we calculate under current plan benefits and contributions.[42] And the reforms have been substantial for most plans in our sample.

41. A more complete analysis of the reforms and changes in contribution level would run the counterfactual starting in 2007, so as to reflect the assets and liabilities that would have prevailed under the counterfactual. Such an analysis, though, is infeasible as our cash flow projection methodology is based on fiscal year 2017 and therefore lacks the ability to perform counterfactual exercises before fiscal 2017.

42. The effect of the benefit changes vary somewhat by stabilization exercise and asset return.

The size of the changes made by the poorly funded plans have been so large as to make those plans among the healthiest when it comes to the stabilization exercises. Plans that made the largest changes in contributions since 2007 and the biggest reforms to their benefits are currently contributing more than enough to stabilize their debt, even at a 0.5 percent rate of return in many cases (compare figure 11 to online appendix tables A3 and A6). To the extent these huge increases in contributions have come at the expense of taxpayer services or higher taxes, it is reasonable to question whether they have been too large. Such an analysis, though, is beyond the scope of the current paper.

VI. Stochastic Analysis

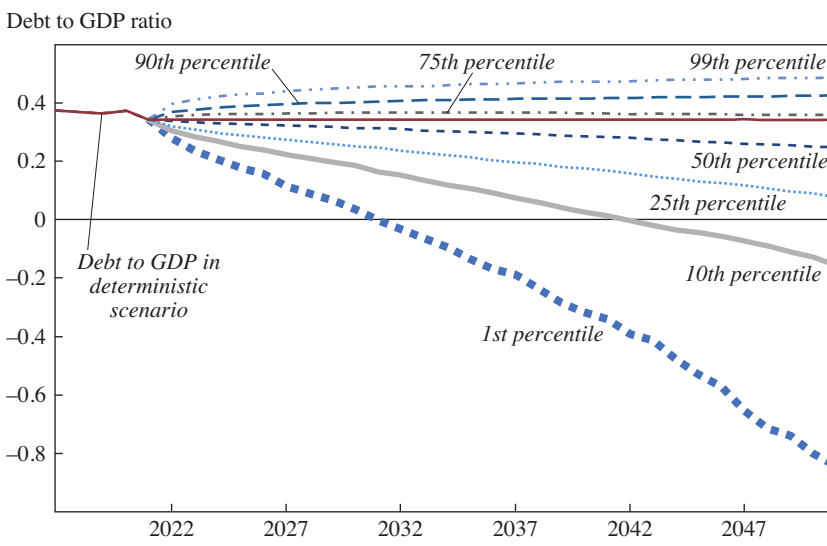
Our approach to asset return uncertainty thus far has been to present results under multiple rates of return, including returns which are lower than the expected rate to account for the costs of risk. An alternative method to explore uncertainty is to calculate the distribution of implicit debt and assets using a stochastic analysis that draws from a distribution of asset returns. For this exercise, plans are assumed to adjust their contributions today so as to hit today's level of pension debt in thirty years' time per the deterministic exercises in the middle panel of table 2. Regardless of the deterministic asset return assumption used to set contribution levels, though, *realized* asset returns are drawn annually from a normal distribution, with a mean return of 6.7 percent (nominal) and a standard deviation of 12 percent.[43] We are therefore implicitly assuming plans continue with their current investment policy. Given our assumption for 2.2 percent CPI inflation, the assumed annual nominal mean rate of return in the stochastic exercises equates to our 4.5 percent real rate of return assumption in the deterministic exercises. The discount rate equals the 0.5 percent real risk-free rate in all cases.[44]

Figure 13 shows the distribution of implicit US aggregate pension debt when plans set their contributions so as to bring implicit debt back to today's

43. These are in keeping with distributions from the literature: see, for example, Yin and Boyd (2019) and "Defined Benefit Program," CalSTRS (California State Teachers' Retirement System), https://www.calstrs.com/defined-benefit-program, which are based on the current composition of pension plans asset portfolios.

44. When assets go negative plans are assumed to issue marketable debt. In the stochastic exercises, the rate of return on this marketable debt is set by the stochastic asset return draws. This is conservative in the sense that a plan may be able to issue debt at a lower mean rate which would improve its fiscal position relative to the results displayed here. That said, in many cases a plan that had exhausted its assets might well be required to pay an elevated rate of return on its debt.

Figure 13. Stochastic Exercise: Debt-to-GDP Percentiles for US Aggregate, Contributions as in Thirty-Year Deterministic Stabilization with 2.5 Percent Real Return on Assets

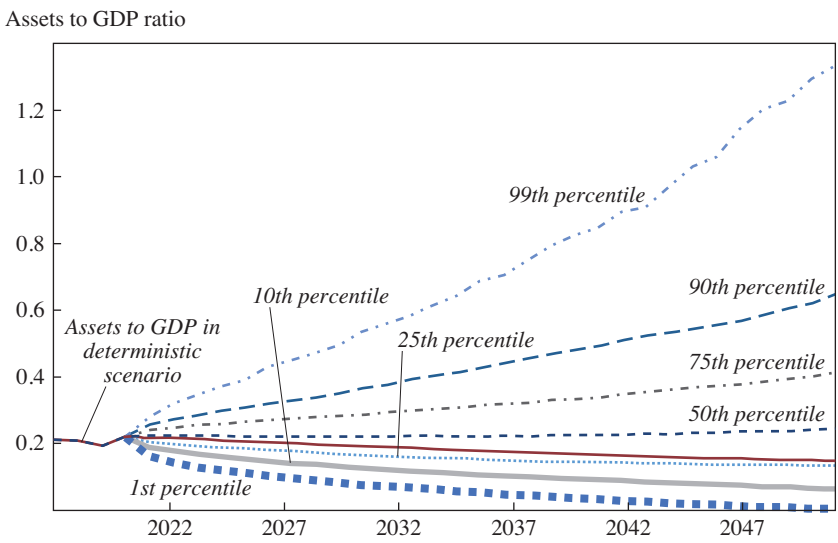


Source: Authors' calculations.

level in thirty years under the deterministic 2.5 percent real rate of return (i.e., contributions increase by 6 percent of payroll as shown on table 2). In this stochastic exercise—in which actual returns average 4.5 percent—pension debt in year 30 falls *below* the starting level of 34 percent of GDP 70 percent of the time; the debt is below 40 percent of GDP 85 percent of the time. But 1 percent of the time, the debt in year 30 rises to above 48 percent of GDP.

If the plans set contributions based on a deterministic real return of 4.5 percent—and therefore lower contributions by 3 percent of payroll so as to have debt back to today's level in thirty years in expectation—the outcomes are less sanguine. Only 37 percent of the time is the debt-to-GDP ratio in year 30 less than the 34 percent starting point; 20 percent of the time it is more than 50 percent of GDP (online appendix figure A7). On the other hand, if the plans base their contributions on the 0.5 percent real rate of return, the median debt at year 30 is 12 percent of GDP; debt is below the 34 percent of GDP starting point 93 percent of the time, and below 42 percent of GDP 99 percent of the time (online appendix

Figure 14. Stochastic Exercise: Assets-to GDP Percentiles for US Aggregate, Contributions as in Thirty-Year Deterministic Stabilization with 2.5 Percent Real Return on Assets



Source: Authors' calculations.

figure A8). Online appendix table A10 presents the distribution of debt in year 30 for this exercise for each plan.

Throughout the paper our focus has been on implicit debt, rather than assets. As we noted above, in the long run, stability of implicit debt implies stability of assets, but nothing in our exercises specifies that those assets be positive. While it might seem intuitive that preventing assets from exhausting is important for plan sustainability—and politically that may be the case—there is little economic difference between small positive or small negative assets (i.e., issuing small amounts of marketable debt to cover liabilities).

Nonetheless, there will likely be political implications to running down pension assets, and so a plan might want to take steps to avoid it. According to our stochastic exercises, and as presented in figure 14, if plans increase contributions to stabilize over thirty years based on a 2.5 percent rate of return, assets in year 30 are positive 99 percent of the time. If they count on a 4.5 percent return, and so contribute less, assets in year 30 turn negative 23 percent of the time, while if they count on a 0.5 percent rate of return, assets in year 30 are positive more than 99 percent of the time. (Online

appendix figures A9 and A10 show the national results under these rate of return assumptions; online appendix table A10 shows the distribution of assets in year 30 for each plan.)

Overall, these stochastic exercises provide a way of linking changes in contribution rates to outcomes. When plans assume a 2.5 percent real rate of return, they face a small possibility of much higher debt and asset depletion over time, even though in the majority of cases, their outcomes will be better than those assumed in the deterministic case. How to assess these various outcomes comes down to understanding how costly risk may or may not be for governments.

The stochastic exercises here are stylized and intended to provide an illustration of the risk around our deterministic debt stabilization paths. A more complete stochastic analysis would simulate based on the specific assets held by each individual plan, consider return distributions other than a normal, and so on. And, of course, a more complete analysis would recognize that many aspects of the pension system and government revenues and expenditures are subject to uncertainty—including wage growth, employment growth, mortality, and tenure. It would also take account of the covariances between sources of uncertainty and include an analysis of the states of the world, in particular levels of marginal utility, in which good and bad outcomes occur. We leave all these considerations to future work.

VII. Conclusion

We find that pension benefit payments in the United States, as a share of the economy, are currently near their peak level and will remain there for the next two decades. Thereafter, the reforms instituted by many plans will gradually cause benefit cash flows to decline significantly. This is an important finding in terms of the fiscal stability of these plans over the longer term as it indicates that the cash flow pressure of these plans will eventually ease. Our results suggest that, under conservative discounting of liabilities and moderate asset return assumptions in aggregate, pension debt can be stabilized with relatively moderate fiscal adjustments. Of course, stabilization costs are higher if asset returns are lower. There is also significant heterogeneity with some plans being far from stable across a range of asset return assumptions. Finally, in aggregate there appears to be only limited advantage to beginning the stabilization process now versus a decade in the future; neither the level at which debt stabilizes as a share of the economy nor the contribution increases needed to achieve stabilization increase much when the start of the stabilization process is pushed a bit farther into the future.

An important limitation to our work is its focus on pension plans in isolation from the broader context of state and local governments. For instance, we implicitly assume that these governments are able to reap the fiscal benefits of pension reforms. However, as employers, state and local governments operate in a competitive labor market; reduction in pension benefits may result in the need to boost other forms of compensation, reducing the fiscal savings from the reforms. Our long-run stabilization scenarios provide another example. In this scenario, governments smooth through the period of peak pension cash flow demand by drawing down assets. Rating agencies might respond to this asset drawdown by lowering credit ratings and we fail to account for the higher borrowing costs for marketable debt that might result. More broadly, the various stabilization paths we explore would ideally be examined through the lens of a cost-benefit analysis incorporating the full policy objectives of these governments. For example, by reducing pension funding governments may be able to increase investments in education and infrastructure. These investments may then yield social returns in the future and also provide fiscal benefits in the form of increased tax revenue. On the other hand, these deficits may carry fiscal costs in the future. We leave these broader considerations for future work.

ACKNOWLEDGMENTS Lorena Hernandez Barcena, Jeffrey Cheng, and Manny Prunty provided excellent research assistance. We are particularly grateful to Lorena, who picked up this project midway through and got up to speed incredibly quickly. We thank Thomas Aaron, Bob Costrell, Ngoc Dao, Greg Duffee, Trevor S. Gallen, Tracy Gordon, Erald Kolasi, Adam Looney, Martin Luby, Robert L. McDonald, Josh McGee, Therese McGuire, Derek Messacar, Jim Poterba, Jonathan Pycroft, Laura Quinby, Kim Rueben, Brian Septon, Juan Carlos Suarez Serrato, David Stemerman, and Tom Terry for helpful comments and suggestions. We particularly thank Jean-Pierre Aubry of the Center for Retirement Research at Boston College, our *BPEA* editor Jan Eberly, and our *BPEA* discussants Deborah Lucas and Josh Rauh.

References

Aubry, Jean-Pierre, and Caroline V. Crawford. 2017. "State and Local Pension Reform since the Financial Crisis." Brief 54. Chestnut Hill, Mass.: Center for Retirement Research at Boston College.

Bagchi, Sutirtha. 2019. "The Effects of Political Competition on the Generosity of Public-Sector Pension Plans." *Journal of Economic Behavior and Organization* 164:439–68.

Blanchard, Olivier. 2019. "Public Debt and Low Interest Rates." *American Economic Review* 109, no. 4: 1197–229.

Bohn, Henning. 2011. "Should Public Retirement Plans Be Fully Funded?" *Journal of Pension Economics and Finance* 10, no. 2: 195–219.

Boyd, Donald J., Gang Chen, and Yimeng Yin. 2019. "Public Pension Risk-Sharing Mechanisms and Their Potential Impacts." Paper prepared for the Municipal Finance Conference, Hutchins Center on Fiscal and Monetary Policy, Brookings Institution, Washington, July 15.

Boyd, Donald J., and Yimeng Yin. 2016a. "How Public Pension Plan Demographic Characteristics Affect Funding and Contribution Risk." Albany, N.Y.: Nelson A. Rockefeller Institute of Government.

Boyd, Donald J., and Yimeng Yin. 2016b. "Public Pension Funding Practices." Albany, N.Y.: Nelson A. Rockefeller Institute of Government.

Boyd, Donald J., and Yimeng Yin. 2017. "Investment Return Volatility and the Los Angeles Fire and Police Pension Plan." Albany, N.Y.: Nelson A. Rockefeller Institute of Government.

Brown, Jeffrey R., Robert Clark, and Joshua Rauh. 2011. "The Economics of State and Local Pensions." *Journal of Pension Economics and Finance* 10 (2): 161–72.

Brown, Jeffrey R., and David Wilcox. 2009. "Discounting State and Local Pension Liabilities." *American Economic Review* 99, no. 2: 538–42.

Bruno, Robert, Amanda Kass, and David Merriman. 2019. *A "Pension Crisis" Mentality Won't Help: Thinking Differently about Illinois' Retirement Systems*. Chicago: Government Finance Research Center, University of Illinois at Chicago; Urbana, Ill.: University of Illinois Institute of Government and Public Affairs.

CBO (Congressional Budget Office). 2011. "The Underfunding of State and Local Pension Plans." Washington. https://www.cbo.gov/sites/default/files/112th-congress-2011-2012/reports/05-04-pensions.pdf.

CBO (Congressional Budget Office). 2019. "The 2019 Long-Term Budget Outlook." Washington. https://www.cbo.gov/publication/55331.

CBO (Congressional Budget Office). 2020. "The 2020 Long-Term Budget Outlook." Washington. https://www.cbo.gov/publication/56598.

CBO (Congressional Budget Office). 2021. "The 2021 Long-Term Budget Outlook." Washington. https://www.cbo.gov/publication/56977.

Cembalest, Michael. 2018. "The ARC and Covenants 4.0." Eye on the Market, JPMorgan Private Bank. https://privatebank.jpmorgan.com/content/dam/jpm-wm-aem/global/pb/en/insights/eye-on-the-market/the-arc-and-the-covenants-4.pdf.

Costrell, Robert M., and Josh McGee. 2020. "Sins of the Past, Present, and Future: Alternative Pension Funding Policies." Paper prepared for the Municipal Finance Conference, Hutchins Center on Fiscal and Monetary Policy, Brookings Institution, Washington, July 13–14.

D'Arcy, Stephen P., James H. Dulebohn, and Pyungsuk Oh. 1999. "Optimal Funding of State Employee Pension Systems." *Journal of Risk and Insurance* 66, no. 3: 345–80.

Duarte, Fernando, and Carlo Rosa. 2015. "The Equity Risk Premium: A Review of Models." *Economic Policy Review* 2: 39–57.

Falkenheim, Michael. 2021. "Government Risk Taking under Market Imperfections." Working Paper. Washington: Congressional Budget Office.

Feldstein, Martin. 1974. "Social Security, Induced Retirement, and Aggregate Capital Accumulation." *Journal of Political Economy* 82, no. 5: 905–26.

Feldstein, Martin, and Jeffrey B. Liebman. 2002. "Social Security." In *Handbook of Public Economics, Volume 4*, edited by Alan J. Auerbach and Martin Feldstein. Amsterdam: North-Holland.

Fitzpatrick, Maria D., and Gopi Shah Goda. 2020. "The Prevalence of COLA Adjustments in Public Sector Retirement Plans." Working Paper. Cambridge, Mass.: National Bureau of Economic Research.

Glaeser, Edward L., and Giacomo A. M. Ponzetto. 2014. "Shrouded Costs of Government: The Political Economy of State and Local Public Pensions." *Journal of Public Economics* 116: 89–105.

Government Accountability Office (GAO). 2008. *State and Local Government Retiree Benefits: Current Funded Status of Pension and Health Benefits*. Washington: Author.

Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright. 2008. "The TIPS Yield Curve and Inflation Compensation." Washington: Board of Governors of the Federal Reserve System.

Haubrich, Joseph G., and Sara Millington. 2014. "PCE and CPI Inflation: What's the Difference?" *Economic Trends*, Federal Reserve Bank of Cleveland, May 6–8.

Hoops, M., P. A. Smith, and I. Stefanescu. 2016. "State and Local Pension Funding in the Enhanced Financial Accounts." Washington: Board of Governors of the Federal Reserve System.

Kamin, David. 2013. "Risky Returns: Accounting for Risk in the Federal Budget." *Indiana Law Journal* 88, no. 2: 723–72.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2012. "The Aggregate Demand for Treasury Debt." *Journal of Political Economy* 120, no. 2: 233–67. https://doi.org/10.1086/666526.

Lenze, David. 2013. "State and Local Government Defined Benefit Pension Plans: Estimates of Liabilities and Employer Normal Costs by State, 2000–2011." Working Paper. Suitland, Md.: Bureau of Economic Analysis.

Lucas, Deborah J. 2012. "Valuation of Government Policies and Projects." *Annual Review of Financial Economics* 4, no. 1: 39–58.

Lucas, Deborah J. 2017. "Towards Fair Value Accounting for Public Pensions: The Case for Delinking Disclosure and Funding Requirements." Prepared

for Harvard Kennedy School Conference, Gathering Storm: The Risks of State Pension Underfunding, October 19–20. Citations are to an unpublished 2020 version.

Lucas, Deborah J., and Marvin Phaup. 2008. "Reforming Credit Reform." *Public Budgeting and Finance* 28, no. 4: 90–110.

Lucas, Deborah J., and Stephen Zeldes. 2009. "How Should Public Pension Plans Invest?" *American Economic Review: Papers and Proceedings* 99, no. 2: 527–32.

Lutz, Byron, and Louise Sheiner. 2014. "The Fiscal Stress arising from State and Local Retiree Health Obligations." *Journal of Health Economics* 38:130–46.

Marron, Donald B. 2014. *A Better Way to Budget for Federal Lending Programs*. Washington: Urban Institute.

Mehra, Rajnish, and Edward C. Prescott. 2003. "The Equity Premium in Retrospect." In *Handbook of the Economics of Finance, Volume 1B*, edited by G. M. Constantinides, M. Harris, and R. M. Stulz. Amsterdam: North-Holland.

Mennis, Greg, Susan Banta, and David Draine. 2018. "Assessing the Risk of Fiscal Distress for Public Pensions: State Stress Test Analysis." Working Paper. Cambridge, Mass.: Mossavar-Rahmani Center for Business and Government, Harvard Kennedy School.

Munnell, Alicia H., Jean-Pierre Aubry, and Josh Hurwitz. 2013. "How Sensitive Is Public Pension Funding to Investment Returns?" Brief 34. Chestnut Hill, Mass.: Center for Retirement Research at Boston College.

Novy-Marx, Robert, and Joshua D. Rauh. 2009. "The Liabilities and Risks of State-Sponsored Pension Plans." *Journal of Economic Perspectives* 23, no. 4: 191–210.

Novy-Marx, Robert, and Joshua D. Rauh. 2011. "Public Pension Promises: How Big Are They and What Are They Worth?" *Journal of Finance* 66: 1211–49.

Novy-Marx, Robert, and Joshua D. Rauh. 2014a. "The Revenue Demands of Public Employee Pension Promises." *American Economic Journal: Economic Policy* 6, no. 1: 193–229.

Novy-Marx, Robert, and Joshua D. Rauh. 2014b. "Linking Benefits to Investment Performance in US Public Pension Systems." *Journal of Public Economics* 116:47–61.

Rauh, Joshua D. 2017. *Hidden Debt, Hidden Deficits: 2017 Edition*. Stanford, Calif.: Hoover Institution.

Samuelson, Paul A. 1958. "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money." *Journal of Political Economy* 66, no. 6: 467–82.

Sastry, Parinitha, and Louise Sheiner. 2015. "Credit Scoring and Scoring of Risk." Paper prepared for The Congressional Budget Office at 40, Hutchins Center on Fiscal and Monetary Policy, Brookings Institution, Washington, May 27.

Sheiner, Louise. 2018. "The Long-Term Impact of Aging on the Federal Budget." Working Paper 40. Washington: Hutchins Center on Fiscal and Monetary Policy at Brookings.

Sheiner, Louise. 2021. "What Debt Crisis? Time to Stop Worrying and Spend What's Needed to Fix the Economy." *Milken Institute Review*, April 26. https://www.milkenreview.org/articles/what-debt-crisis?IssueID=40.

Shoag, Daniel, and James Farrell. 2017. "Risky Choices: Simulating Public Pension Funding Stress with Realistic Shocks." Working Paper 37. Washington: Hutchins Center on Fiscal and Monetary Policy at Brookings.

Society of Actuaries. 2014. *Blue Ribbon Panel on Public Pension Plan Funding*. Schaumberg, Ill.: Author.

United States Congress. 1978. *Pension Task Force Report on Public Employee Retirement Systems*. Washington: US Government Printing Office.

United States Congress Joint Economic Committee. 2012. "The Pending State Pension Crisis." Blog post, September 26, JEC Republicans. https://www.jec.senate.gov/public/index.cfm/republicans/2012/9/the-pending-state-pensions-crisis.

Winkelvoss, Howard E. 1993. *Pension Mathematics with Numerical Illustrations,* 2nd edition. Philadelphia: University of Pennsylvania Press.

Yin, Yimeng, and Donald J. Boyd. 2019. "Analyzing the Interplay between Public Pension Finances and Government Finances: Lessons from Linking an Economic Model to a Pension Fund Model." Albany, N.Y.: Nelson A. Rockefeller Institute of Government.

Comments and Discussion

COMMENT BY

DEBORAH LUCAS As much of my discussion is critical of the approach taken in this paper, I want to start on a positive note and sincerely applaud the authors for their contributions to the discussion of pension fund sustainability. Their careful projections of the cash flows associated with future state and local pension liabilities are particularly valuable. It's said that imitation is the sincerest form of flattery, and so I'll note that we have a project underway at the MIT Golub Center for Finance and Policy to project the probability distribution of future cash flows and funding levels for a larger universe of public plans. I can attest to how much work it takes to create each new data point.

I also want to highlight a few of the authors' conclusions where I am in agreement. The first is that there is unlikely to be a widespread crisis caused by state and local pension plans being unable to meet their liabilities, at least absent a major meltdown in the stock market. Even if some plans run out of funds much sooner than suggested by the analysis in this paper (and I believe some will), exhaustion dates are likely to be staggered over time. Hence, it is incorrect to view pervasive underfunding as a harbinger of an impending national crisis. Second, I concur with the lack of an imperative to quickly or fully close funding gaps, although as explained below, the theoretical and practical considerations that lead me to those conclusions are distinctly different from the ones suggested in this paper.

However, the authors go much further in challenging the conventional wisdom. They argue that not only is there no looming nationwide crisis but that there is no reason to be concerned about the vast majority of individual plans. They further provocatively suggest that it might well be desirable

to divert funds earmarked for pension funding to other uses and that such policies are likely to be sustainable.

Those sanguine conclusions are the result of several questionable modeling choices that include a trip down the dangerous rabbit hole of $r - g$ accounting; abstracting from uncertainty; and treating the information in financial market prices as irrelevant to public finance. I believe that any analysis that avoided those shortcomings would conclude that there is a strong link between funding levels and sustainability. The remainder of my comments elaborate on the following observations.

The authors' calculations rest on accounting identities that critically depend on the assumed return on assets, r, and the growth rate of GDP, g. The authors take these variables to be deterministic and only loosely justify the assumed relation between them. In fact, the restriction that r must be greater than g is a robust implication of any standard, deterministic, general equilibrium model. When growth rates and asset returns are stochastic, a welfare analysis based on comparing average r and average g has no theoretical foundation. Meaningful evaluation of fiscal policy alternatives requires incorporating the effects of uncertainty and its associated costs.

The assumptions of perpetually low interest rates and asset returns, which are necessary to conclude that waiting to fill funding gaps makes little difference, rest on shaky empirical grounds.

Pension plan sustainability is equated in this analysis to projecting a finite long-run debt-to-GDP ratio in a deterministic world. A more conventional view of sustainability is that it requires a plan to be able to withstand the risk of adverse shocks such as lower than expected asset returns or unanticipated population loss over some extended period of time.

Under a definition of sustainability that focuses on the likelihood of asset exhaustion over a specified horizon, increasing funding always improves sustainability because it increases a pension plan's distance to default. A larger number of plans would be classified as unsustainable under this definition.

The authors suggest that increases in pension fund contributions have reduced public investment. It is important to recognize that there is no economic trade-off between the two uses of funds. If there is crowding out, it is a political phenomenon, and the potentially high costs of relaxing pension funding requirements suggest it is probably better addressed by other means.

THEORETICAL FOUNDATIONS In recent years, the (short-term, risk-free) interest rate, r, has been lower than the GDP growth rate, g. This has generated optimism among some commentators that accumulating high levels

of government debt is sustainable. If relatively low interest rates were to persist indefinitely, GDP growth would exceed debt service cost growth, and tolerating high levels of debt-to-GDP would place a small or even shrinking debt service burden on future generations. Put differently, if *g* is permanently greater than *r*, then it is better to borrow and invest in real growth rather than to pay down government debt; there is essentially an arbitrage opportunity.

If policymakers are considering running a high-debt policy based on these observations, it is important to ask first whether it is reasonable to expect interest rates to remain below growth rates indefinitely. The prediction of a standard, deterministic, equilibrium model suggests the answer is no.

The logic that *r* will exceed *g* in a deterministic equilibrium can be illustrated very simply in a two-period setting that can be shown to generalize to an infinite horizon. Assume that aggregate income grows at a constant rate *g*. Agents maximize utility through the choice of how much first-period income to save, *S*, subject to the usual lifetime budget and wealth constraints:

$$\max_{S} U(C_1) + \beta U(C_2),$$

such that $C_1 = Y_1 - S$ and $C_2 = Y_2 + S(1 + r)$. Differentiating with respect to *S* implies the optimality condition:

$$\frac{1}{1+r} = \beta \frac{U'(C_2)}{U'(C_1)} = \beta \frac{U'(C_1(1+g))}{U'(C_1)}.$$

The interest rate, which must satisfy the optimality condition in order to clear the market for borrowing and lending, depends on *g*. For any concave utility function, the market-clearing value of *r* increases in *g*. When $U(C) = ln(C)$, the equation simplifies to $(1 + r) = (1 + g)/\beta$. If $\beta < 1$, then $r > g$ (e.g., $\beta = .98$, $g = 1$ percent implies $r = 3.06$ percent). With log utility, $r - g$ is roughly the subjective rate of time preference.[1]

What if interest rates and growth rates are uncertain? A few years before Samuelson (1958) presented the accounting identity used in this analysis, Arrow and Debreu (1954) introduced the elegant and powerful concept of state prices—the recognition that a dollar in different future states of the world will have a different utility value relative to a dollar today,

1. For constant relative risk aversion (CRRA) preferences, the higher the coefficient of relative risk aversion, the stronger is the preference for consumption smoothing over time, and hence the larger is the positive difference between *r* and *g*.

depending among other things on the future strength of the economy: a claim on resources in bad times is worth more than the same claim in good times. That idea is at the core of modern finance. It explains why the expected return on risky assets like stocks, whose payoffs are positively correlated with aggregate consumption, exceeds the return on safe government debt. These insights about how risk influences value are as critical for assessing the welfare implications of fiscal policy choices as they are for understanding private sector asset returns.

We can now turn back to the relation between r and g in a similar setup to the one described above but incorporating uncertain growth. In that setting, an induced precautionary demand for savings may imply an average risk-free rate that is less than the average growth rate of consumption. However, as just explained, the value today of uncertain future output can no longer be inferred by discounting expected payoffs at the risk-free rate. Hence it is impossible to conclude on the basis of average g exceeding average r that it is desirable for the government to postpone paying debt or to issue more of it in order to invest in growth assets. To do so treats risk premiums as arbitrage opportunities rather than as market-determined compensation for bearing costly aggregate risk.

This relates directly to the authors' calculations, some of which assume a fixed return on assets that is higher than the risk-free rate that affects the future value of liabilities. In this case the analysis has a built-in arbitrage opportunity. Had the authors also taken into account that a pension plan (or the government sponsor) could borrow risk free to invest in plan assets, the accounting identity would show this to be a money machine that makes it costless to eliminate underfunding.

Perhaps anticipating this criticism, the authors also look at a "certainty-equivalent" case where investment assets are assumed to earn the risk-free rate. While the certainty-equivalent methodology can be a convenient tool for inferring the market value of a risky asset, using it here to extrapolate future outcomes is problematic. It is noteworthy that one of the most provocative claims of the analysis—that there isn't much gain to closing funding gaps sooner rather than later—rests on this assumption. That's because when asset returns are low and unfunded liabilities grow at the same low rate, acting sooner has little advantage. Viscusi (2007) points this out in the context of explaining the paradoxes that arise when the social discount rate is taken to be zero. However, if the goal is to forecast average outcomes, simple algebra implies that assuming typical plan portfolio choices, higher funding levels would reduce the average debt service burden on future generations.

If the reader is wondering what I am recommending to replace the authors' assumptions about interest rates and asset returns, it is to do a stochastic simulation that takes into account the higher expected return on assets and also the higher risk. This is discussed in the context of sustainability below.

The insights of state pricing are also critical for evaluating costs to future generations and whether they will be equitably distributed. A pension plan is more likely to run out of money when it is highly underfunded to begin with and during a sustained economic downturn. In order to cover a funding shortfall, the sponsoring government may be forced to cut other spending. Those spending cuts, which are especially costly because they occur when the economy is depressed, will be largely borne by the unlucky current residents of that locale. Hence, under realistic assumptions about government behavior in response to cash shortfalls, the costs of underfunding will fall disproportionately on generations that are least able to afford it. This is in sharp contrast to the assumption of complete smoothing of costs across generations through the mechanism of debt financing, and the reason why I think that sort of analysis cannot be used to draw conclusions about the welfare consequences or distributional impacts of underfunding.[2]

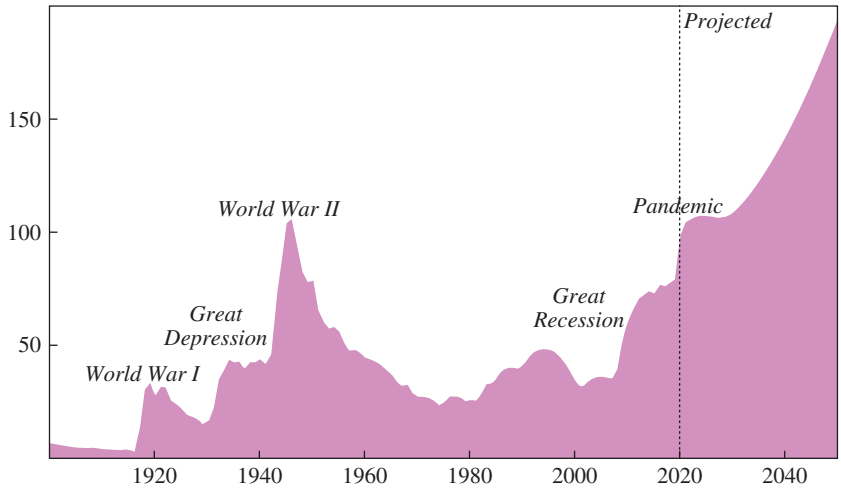
On the issue of fairness between generations, spreading underfunding evenly over future generations may not be perceived as equitable. In fact, a standard argument for fully prefunding accruing benefits is on equity grounds: the primary beneficiaries of current public services then bear the full cost. There is a stronger case for spreading legacy underfunding across multiple generations, but there are nuances involving relative wealth and incentive effects that may weigh against it.

MORE ON PARAMETER CHOICES As the calculations and conclusions of the paper are difficult to interpret because of the foundational issues just discussed, I only have a few comments on specific parameter choices.

Unlike most of the variables, which are justified with reference to historical data, the risk-free rate r is loosely based on the low-rate conditions of 2020. Historical data suggest that r, and also $r - g$, has been lower recently than it has been on average. It is not clear why the authors choose to deviate from the standard practice of equating fixed parameter values to historical

2. Citing an unpublished CBO paper, the authors also claim that "to the extent the risk premium reflects business cycle risk, the government can lower that risk by spreading it across future generations." Even if the government redistributes risk within the population so as to reduce the welfare costs of business cycles, the equilibrium equity premium reflects the cost of the risk that remains. That premium is relevant to assessing the cost of risk associated with the government's fiscal policies.

Figure 1. Federal Debt Held by the Public, 1900 to 2050



Source: CBO (2020).

averages when it comes to *r*, but that choice is important for the conclusion that there is little gain from closing the funding gap more quickly.

A minor quibble is that the authors justify some parameter choices on the basis that CBO makes similar assumptions. CBO's assumptions are not a natural reference point because it makes current law projections used in baseline exercises, not statistical forecasts. Over long horizons the difference between the two can be significant.

SUSTAINABILITY REVISITED Whether or not the ratio of long-run debt-to-GDP can be stabilized (i.e., it remains finite as time is taken to infinity) in a deterministic world seems largely orthogonal to whether public pension funds are sustainable for at least two reasons. First, a stable debt ratio does not imply a manageable debt service burden. Second and more fundamentally, sustainability requires resilience to adverse events, and that can only be assessed in a stochastic framework.

With regard to level effects, a stable but high debt-to-GDP ratio would probably be viewed as unsustainable, or at least very undesirable, when it entails high debt service costs paid for with distortionary taxes. In assessing the burden, the cost of servicing federal and other public debt also has to be factored in. As shown in figure 1, the federal debt-to-GDP ratio is projected to reach unprecedented levels, doubling over the next thirty years and showing no sign of stabilizing. It is hard to imagine that adding to

those projected debt ratios by deliberately increasing pension underfunding would be considered sustainable or wise.

Focusing on the level rather than the stability of the debt ratio also makes the risk of future interest rate hikes more salient to the question of sustainability. As the debt ratio rises, the debt becomes riskier and investors require a higher credit spread to buy it, pushing up the cost of debt service. In extreme cases, it will be impossible to roll over the debt and public services or pension benefits will have to be cut. Notice that the properly measured prospective social cost of such adverse outcomes exceeds the average increase in interest payments and services lost because the costs tend to be realized when the economy is weak. The policy of tolerating higher debt ratios transfers costly market risk to future generations. This is another example of where the policy-induced cost of market risk remains hidden when analysts think only in terms of average cash flows.

Long-run stability is not a sufficient condition for sustainability. An underfunded pension plan, despite being characterized as stable in the long run, may run out of available resources to pay promised benefits due. A pension plan with a high chance of becoming insolvent is not sustainable as the term is commonly understood. There's an analogy in the Social Security system. Whatever is projected about future taxes and benefits, if the trust fund hits zero, there is no budget authority to pay current benefits if they exceed current payroll tax revenues; the system is unsustainable without political action. This is true even though the trust fund is an accounting mechanism and not a repository of financial assets. In the same way, if a public pension plan's assets fall to zero and current contributions are insufficient to cover current benefits, it will trigger a crisis that will require legislative or executive action to address.

I would therefore propose an alternative definition of sustainability: a fund is sustainable if it can meet its contractual obligations with sufficiently high probability over a specified period of time. Operationalizing this definition would require choosing a threshold probability and time horizon, for example, a system could be classified as sustainable if there is a 95 percent probability that assets will not be exhausted at any point over the next two decades. This alternative definition recognizes that assessing sustainability requires modeling funding levels as stochastic. It also requires computing the time path of the distribution of funding outcomes, not just its average.

Under that definition, higher funding levels unambiguously improve sustainability, including in low-rate environments. While many sources of uncertainty affect funding levels, arguably the largest source of year-to-year volatility is a plan's risky asset holdings. Assuming that asset returns are

normally distributed (and not too highly correlated with other determinants of funding levels), the formula for the probability of first passage times for a normally distributed variable implies that the probability of assets going to zero over a specified horizon is monotonically decreasing in the level of assets. This is analogous to the concept of distance to default used in standard analyses of default risk for businesses. Closing a funding gap more quickly also always improves sustainability under this definition.

For a given funding level, incorporating more realistic assumptions about expected asset returns and volatility has partially offsetting effects on sustainability. A higher average return increases sustainability while volatility reduces it. The prior literature on this that the authors cite, and my own recent work (Lucas and Smith 2020) on whether a collective defined contribution system can deliver a fairly safe and adequate benefit without boosting contribution rates (we find it can't), suggests that taking stochastic returns into account would cause more plans to be classified as unsustainable, again under a definition based on the likelihood of asset exhaustion. The authors have the data to simulate the probability distribution of exhaustion dates for each of the plans in their database when asset returns are stochastic, and it would be very informative to see those results.

FURTHER OBSERVATIONS ON THE COSTS AND BENEFITS OF PREFUNDING As noted at the beginning of my remarks, I agree with the authors that there is no imperative to quickly or fully close funding gaps, but for distinctly different reasons than the ones they emphasize.

The authors suggest that a benefit of relaxing pension funding requirements is to avoid crowding out public investments such as in infrastructure and education. However, because most state and local governments have access to capital markets or bank financing, from a purely economic perspective there is no reason for the two goals to compete. Governments can borrow to turn unfunded pension liabilities into funded ones, or to pay for investments, or both. Total pension liabilities might rightly influence the perceived affordability of other spending plans, but fundamental affordability should be invariant to pension funding status. That reasoning applies the logic of Modigliani and Miller's (1958) famous capital structure irrelevance theorem to public pension plans, and its implications are explored in more detail in Lucas (2017). In practice, crowding out may occur, but if so, it is a political phenomenon that can be addressed in other ways. For example, a state legislature could weaken its self-imposed balanced budget requirement.

Relatedly, the main reason that I think a narrow focus on full funding is misplaced is the fact that funding gaps can be reduced using borrowed funds, and hence increased funding may have no impact on current or future

fiscal policies. The authors make the same point when they say, "More broadly, governments typically hold debt, and unfunded pension liabilities are simply a form of (implicit) debt." Although we agree on that fact, it doesn't alleviate my concerns about the consequences of high consolidated debt levels. Evaluating the sustainability of the public finances of a city or state requires also taking into account other current and projected liabilities, and the possible paths of future spending and revenues. This is essentially what municipal bond rating agencies do.

There are other legitimate reasons one might favor higher funding levels beyond concerns about the possibility of a funding crisis or eventual limits on the capacity to borrow. Those include imposing fiscal discipline, making opaque liabilities more transparent and salient to the public, discouraging governments from making unsustainable benefit promises, and ensuring that the cost of services falls on the current beneficiaries and not on future generations.

REFERENCES FOR THE LUCAS COMMENT

Arrow, K. J., and G. Debreu. 1954. "Existence of an Equilibrium for a Competitive Economy." *Econometrica* 22, no. 3: 265–90.

CBO (Congressional Budget Office). 2020. "The 2020 Long-Term Budget Outlook." Washington. https://www.cbo.gov/publication/56516.

Lucas, Deborah J. 2017. "Towards Fair Value Accounting for Public Pensions: The Case for Delinking Disclosure and Funding Requirements." Prepared for Harvard Kennedy School Conference, Gathering Storm: The Risks of State Pension Underfunding, October 19–20.

Lucas, Deborah J., and Daniel Smith. 2020. "How Much Can Collective Defined Contribution Plans Improve Risk-Sharing?" *Journal of Investment Management* 18, no. 3.

Modigliani, F., and M. H. Miller. 1958. "The Cost of Capital, Corporation Finance, and the Theory of Investment." *American Economic Review* 48, no. 3: 261–97.

Samuelson, Paul A. 1958. "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money." *Journal of Political Economy* 66, no. 6: 467–82.

Viscusi, W. Kip. 2007. "Rational Discounting for Regulatory Analysis." *University of Chicago Law Review* 74, no. 1: 209–46.

COMMENT BY

JOSHUA RAUH The question of the sustainability of pension promises is a good one. It is well known that most state and local government pension promises are underfunded, but what kind of adjustments must

governments make to their contributions in order to prevent these unfunded liabilities from causing fiscal instability and debt crises?

If the accounting for pension promises were based on measuring liabilities at their actual financial value rather than at an artificially low value based on expected return, then one could very well entertain arguments that the optimal pension funding ratio is not 100 percent, just as the optimal amount of public debt is unlikely to be zero. The authors are correct to point out that the goal of moving from the currently underfunded status of pension promises to full prefunding is a goal that would be analogous to paying down government debt, as opposed to stabilizing it as a percentage of GDP.

There is an important caveat here, however, which is that when pension sponsors can measure unfunded liabilities using expected returns on plan assets, they may be incentivized to ignore risk in the setting of pension funding policy. While I agree in principle that achieving a stable ratio of properly measured government debt to GDP is a sufficient goal for financial stability, the narrower goal of targeting a stable value of unfunded pension liabilities as a share of GDP may not be sufficient for financial stability—especially when sponsors' measurement of those liabilities biases them toward taking risk in order to meet expected returns dictated by a political process. Even though the authors' analysis targets a debt ratio at a correctly measured discount rate, pension funds still must set an expected rate of return and an asset allocation that targets that rate of return.

That setup by itself may introduce instability. Indeed, partially funded or even fully unfunded PAYGO can be sustainable under the right conditions. However, the potential for instability arises in this context because of the possibility that a pension system would have to start paying benefits out of current resources, requiring suddenly much higher draws on current resources than under a 7.25 percent return assumption. The authors effectively eliminate this possible instability by assuming plans can borrow after exhausting their assets. It is the concern that municipal credit markets might view this risk as substantial that creates instability, since that would then lead to large increases in borrowing rates and challenges to the ability of municipalities to access credit markets.

One of the features of the analysis in this paper is that the discount rate and the expected return play separate roles. In the context of the simulation-based approach adopted by the authors later in the paper, this is potentially fruitful, as of course the discount rate for a fixed pension promise and the expected return on a risky portfolio of assets should be different. For deterministic analysis, however, it makes little sense to choose rates of return that deviate from the risk-free rate. Based on the analysis by Costrell and

McGee (2020) of the authors' model, the deterministic model seems to have the feature that for any given steady-state value of assets to payroll, the steady-state contribution rate is independent of the discount rate and depends only on the expected return. In fact, what seems to happen is that lower discount rates increase the normal cost but (holding the expected return fixed) allow the difference between expected return and risk-free rate to offset that additional cost.

Under the fundamental theorem of finance, the risk-adjusted expected return on assets is the risk-free rate. So if one has to pick only one point to represent the distribution, then this is the right one. Under this scenario, pension debt to nominal GDP would rise indefinitely without fiscal adjustments. For these reasons, the most appropriate single parameterization is one with $r = \delta = 0$ percent. The thirty-year TIPS yield is –0.15 percent as of June 14, 2021, so there is no reason to use something higher. A real yield of –0.15 percent is below the lowest real return assumption in table 2 of the paper, but it is close to the 0.5 percent parameterization, so for most of the goals examined in table 2, I think of the correct percentages as a few points higher than those shown in the 0.5 percent return case. Taking that 0.5 percent scenario at face value, the required contribution increase to get implicit debt back to today's levels is 14.7 percent of payroll if they start today. Relative to the current weighted contribution rate of 24 percent in the authors' estimation sample, the 14.7 percent of payroll hike amounts to an increase in contributions by over 60 percent. This is substantial.

Appropriately, the authors have introduced a stochastic analysis that allows them to consider the distribution of outcomes generated by the risk-loaded investment strategies of public pension funds. Here it does make sense to consider this distribution of possible outcomes. As is well known in finance, the likelihood of exceeding the risk-free rate of return is high under standard lognormal return distribution assumptions, and the distribution of outcomes shown by the authors reflects this.

There are a number of points that must be considered in interpreting the paper's simulation analysis. First, it must be recognized that the states of the world where bad return outcomes occur are high marginal utility states. This is precisely the reason why the stock market has an expected return that is higher than the risk-free rate. Under risk-neutral probabilities, which put more weight on bad states of the world, the expected return remains the risk-free rate.

Second, for their main stochastic simulation (figure 13 in the paper), the authors are assuming contribution rates based on a strategy in which contributions are set under a 2.5 percent real return assumption, while actual

returns on assets in the simulation average a 4.5 percent real return (e.g., 6.7 percent nominal and 2.2 percent inflation). The authors' caveat that if the plans set contributions based on a deterministic real rate of return of 4.5 percent, decreasing contributions by 3 percent instead of increasing them by 7 percent, "the outcomes are less sanguine." But of course, public pension plans are setting their contribution rates today based on their expected rates of return, not based on lower rates of return. The introduction of any strategy like targeting a debt level in thirty years equal to that of today would be unlikely to move public pension systems toward more conservative return assumptions in setting their contribution rates. Therefore, this less sanguine distribution of outcomes, shown in online appendix figure A7, seems a more appropriate illustration of the likely impact of these policies.

Third, the risk-taking necessary to target high returns exposes pension systems to significant volatility—not only a distribution of long-term return outcomes but also a range of possible paths that lead to those outcomes, including some that could lead to near-term insolvency. The issue is not only the possibility of ending the thirty-year time period with negative assets but of having to avoid crossing over to negative assets by conducting large debt issues in the interim.

This last problem arises as well in the main deterministic analysis. Waiting thirty years leads to a smaller required adjustment. Fifteen plans (37.5 percent of those in the sample) are insolvent before thirty years, under current contribution rates (online appendix table A2). The authors rely on the idea that plans can issue debt to "smooth through the period of peak pension cash flow demand," yet add any volatility and there is a chance that even more could become insolvent. It seems critical that benefit growth really would slow dramatically. It also seems that waiting would stabilize the funding ratio at a much lower level. The intuition for why this is ultimately less costly seems to be that GDP is higher (and normal costs are slightly lower) at the time when the adjustment would start.

This gets to the other main driver of the authors' results for the goal of stabilizing implicit debt to GDP under the (nearly) risk-free rate scenario, and that is that the discount rate (and rate of return) is less than the rate of economic growth ($\delta = r < g$). There are many reasons why such an assumption may not hold. Not least of these is that the growth rate g is the growth rate of public payrolls, which cities and states might well lower if they need to increase contributions.

A further consequence of this modeling is that the authors' recommended contribution rates are actually well below the normal cost rate (Costrell and McGee 2020). An online appendix table provided by the authors

(table A9) allows the reader to see this. Baseline normal costs measured at the 0.5 percent rate are 51.3 percent of payroll in 2017 and 47.5 percent in 2047. This highlights the fact that the goal of stabilizing the implicit debt-to-GDP ratio involves the recommendation of contributing less than the normal cost. The authors do not seem too worried about this, but to me it highlights the fundamental dynamics in the paper: due to the assumptions about payroll growth and returns, the authors are concluding that plans can contribute less than the present value of newly accrued benefits and still achieve stability. Since online appendix table A9 shows that normal costs are not really declining by that much over time, this must involve plans bearing substantial financial risk.

In sum, this paper provides a wealth of calculations and scenarios that will be useful to anyone studying this issue. My takeaways from the analysis are somewhat different from the conclusions of the authors: stabilization would require quite substantial contribution increases of 50 percent or more on average, and much more for specific plans; scenarios that involve waiting to address the issue or goals of stabilizing debt to GDP depend heavily on the $\delta < r < g$ assumption, which may not be appropriate; state pricing requires us to put more weight on bad states of the world when assessing scenario analyses; and the volatile paths that assets might take require future modeling to address the possibility of interim insolvency if pension systems are going to be content with goals less aggressive than paying down unfunded liabilities.

REFERENCE FOR THE RAUH COMMENT

Costrell, Robert M., and Josh McGee. 2020. "Sins of the Past, Present, and Future: Alternative Pension Funding Policies." Paper prepared for the Municipal Finance Conference, Hutchins Center on Fiscal and Monetary Policy, Brookings Institution, Washington, July 13–14.

GENERAL DISCUSSION Henry Aaron remarked that Deborah Lucas's comments suggested that volatility may make pensions unsustainable by increasing the risk that assets will be exhausted. Aaron questioned whether, taking risks into account, pension managers could take steps to insure against or reduce the likelihood of exhaustion, and he wondered how administratively feasible or effective such insurance would be. He asked if the relative optimism of the paper would be restored if such steps were feasible.

Deborah Lucas responded to Aaron's comments by noting that what is salient from her analysis of the defined benefit contribution plan is that

it found that it was not possible to guarantee people a reasonable benefit using investments in a mix of risky and safe assets and to also have the existing contribution rates.[1] She noted that this implies that if a pension invests in enough risky assets to create the desired benefit levels with the existing contribution levels, the pension is relying on an equity premium with a risk that the fund will run out. Lucas said that if a pension bought insurance, then the cost of the insurance is the difference between the risk-free liability and the risky assets. Governments could pre-commit to diverting revenues intended to pay for other spending to make pension benefits safer, but that would be politically difficult.

Janice Eberly noted that both Lucas and Joshua Rauh recommended adding uncertainty to the analysis in their comments. She wondered whether the work that Lucas had already done changed the authors' view on heterogeneity. Eberly remarked that she could see how plans could be different because of investment profiles and asked if there were other significant sources of variation that would add to that heterogeneity besides the ones the authors had already identified.

Rauh responded to Eberly's question, emphasizing that dealing with heterogeneity requires using a stochastic simulation such as a Monte Carlo simulation. He noted that standard error bounds are not as effective due to the asymmetry in the way that asset returns are typically modeled (such as in the Black-Scholes-Merton lognormal model used in his comments), and that the median outcome is much below the expected return in this type of model.

Eberly clarified that she was asking if adding risk would provide different results for different plans in excess of the heterogeneity from investment profiles. She thought that Rauh's comment about asymmetry might imply that the answer was yes.

In response to Eberly's question, Louise Sheiner commented that the paper took a public finance approach and thought about state and local pensions as government debt. The analysis smoothed through heterogeneity in order to take a broad look, similar to looking at the unified budget instead of the Social Security Trust Fund. She noted that stochastic analysis would be both helpful and relevant to the analysis, but that it was more complex than accounting for heterogeneity in returns. Sheiner observed that there are multiple factors such as wages, demographics, and GDP that affect returns but also have a direct impact on pensions. She also commented that the

1. Deborah Lucas and Daniel Smith, "How Much Can Collective Defined Contribution Plans Improve Risk Sharing?," *Journal of Investment Management* 18, no. 4 (2020).

authors thought about sustainability analysis similarly to work done for the federal government using projected cash flows. She mentioned that since a lot of revenues for the federal government come from capital income, expected values are used. Sheiner said that while there is a lot of uncertainty surrounding this type of analysis, most people seem to be comfortable with this methodology for the federal government. She also pointed out that a state like California has a lot of similarities to the federal government, particularly in the fact that it can respond if something goes poorly.

Caroline Hoxby noted that there is literature that suggests that public sector employees would be willing to accept less than the present value of their future pensions if they got more of it in current income. Hoxby also mentioned that there are political economy reasons why government employees tend to receive compensation in pensions rather than in salaries or present-day income. She asked if the authors had considered this when thinking about stabilizing pensions. David Wessel asked if unfunded liabilities are providing misleading metrics for policymakers, since the unfunded liability approach does not fully incorporate changes in pension benefits for new hires.

Sheiner responded to Hoxby's point by saying that the authors have not considered the literature mentioned, but she noted that plans may be able to lower compensation in the future, since the current competitive labor market equilibrium implies higher wages. She mentioned that interest rates have come down significantly, which has increased the value of defined benefit plans, but that this is not something that will continue going forward. Sheiner also mentioned that approximately 25 percent of state and local government employees are not covered under Social Security and that they would be willing to receive higher compensation instead of Social Security benefits.

In response to Wessel's question on misleading metrics, Sheiner said that the authors believed that the value of the paper was in the similarities of the public pension space to the lowering of rates of retirement twenty years in the future during the Greenspan Commission. From a long-run public perspective, Sheiner noted that smoothing out was necessary, and while assets might dwindle now, there would be relief later, unlike the Baby Boom, when benefits went up and stayed up.

Rauh, in agreement with Sheiner on her response to Wessel, said that he believed that the unfunded liability ratio was not any more or less misleading than the debt-to-GDP ratio. Sheiner noted that the measurement looks at current stock, not flows to project, and that the authors would like to extend this analysis for state and local governments as a whole.

Panel on COVID-19 and Economic Policy after One Year

ANDREW ATKESON
University of California, Los Angeles

Behavior and the Dynamics of Epidemics

ABSTRACT A model of private and public behavior to mitigate disease transmission during the COVID-19 pandemic over the past year in the United States addresses two questions: What dynamics of infections and deaths should we expect to see from a pandemic? What are our options for mitigating the impact of a pandemic on public health? I find that behavior turns what would be a short and extremely sharp epidemic into a long, drawn-out one, with, at best, a modest impact on the long-run death toll from the disease. Absent the development of a technological solution, such as vaccines or life-saving therapeutics, additional public health interventions suffer from rapidly diminishing returns in improving long-run outcomes. In contrast, rapidly implemented nonpharmaceutical interventions, in combination with the rapid development of technological solutions, could have saved nearly 300,000 lives relative to what is now projected as of mid-June 2021 to occur over the long run.

During the first half of the twentieth century, Americans enjoyed tremendous gains in health and life expectancy as infectious diseases were drastically curtailed thanks to major medical advances and significant investments in sanitation and public health. Annual mortality rates from infectious disease in the United States fell by an order of magnitude from nearly 800 per 100,000 in 1900 to under 50 per 100,000 by 1960, in a steady downward trend interrupted, dramatically, by the 1918–1919 influenza epidemic (Armstrong, Conn, and Pinner 1999).[1] But as the HIV/AIDS pandemic made evident, and the COVID-19 pandemic reinforced,

Conflict of Interest Disclosure: The author did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this paper. He is currently not an officer, director, or board member of any organization with an interest in this paper.

1. To place the mortality from COVID-19 in historical perspective, note that in the United States it was roughly 100 per 100,000 in 2020 and may very well reach this level again in 2021. So while mortality from COVID-19 will not reach the levels experienced during the Spanish flu, it will clearly be the most significant short-term increase in mortality from infectious disease in the United States in at least sixty years.

infectious diseases are far from vanquished (Morens and Fauci 2020). In fact, the risk of experiencing another pandemic in the not too distant future is considerable. For example, according to a September 2019 estimate by the president's Council of Economic Advisers, there is a 4 percent probability of an influenza pandemic annually. If such a pandemic were to occur it would, at the high end, cause nearly $4 trillion in economic damage and over half a million deaths (Council of Economic Advisers 2019).

Given that we are likely to see significant outbreaks of infectious disease in the future, this moment, after a year of COVID-19, seems an opportune time to reexamine our models of disease dynamics and the policy options for disease control implied by these models. What dynamics of infections and deaths should we expect to see from a pandemic? What are our options for mitigating the impact of a pandemic on public health? How might this mitigation be done in a manner to reduce the negative impact of a pandemic on the economy? These are questions that will provoke new research in light of worldwide data from this COVID-19 pandemic for years to come.

But, with one year of data on COVID-19 now available, one conclusion seems clear: the endogenous response of both public and private behavior to the prevalence of COVID-19 has transformed this epidemic from what standard epidemiological models predicted to be a short, but exceedingly intense, episode into a drawn-out pandemic that will have an impact on public health and the economy over several years, until, with luck, the technological solutions of vaccination and life-saving therapeutics bring this disease under much greater control worldwide.[2]

In this paper, I use a simple model of our experience with COVID-19 in the United States over the past year to explore how the interaction of disease and behavior changes the dynamics of an epidemic and constrains our options for mitigating the impact of a pandemic on public health absent a technological solution such as vaccines and life-saving therapeutics. Based on this model, I present four conclusions.

First, the behavioral responses that we have seen to COVID-19 over the past year, both private and public, have had a powerful impact in "flattening the curve," reducing peak levels of daily infections and deaths by an order of magnitude relative to predictions of standard epidemiological models. These behavioral responses, however, are forecast to have only a

2. It is clear that the development of vaccines for COVID-19 has been a technological marvel. There also appears to be considerable promise for the development of therapeutics that could substantially reduce the severity of the disease and thus complement vaccines in bringing the pandemic to an end worldwide. See Saelens and Schepens (2021) for a description of such therapeutics.

modest impact in reducing the long-term death toll from COVID-19 relative to predictions of standard epidemiological models in the absence of the development of technological solutions such as vaccines or life-saving therapeutics. Absent such technological solutions, the long-run death toll in the United States would approach 1.25 million over a five-year period, even with the private and public efforts at mitigation that have been undertaken. Thus, without the success of vaccines that we have now experienced here in the United States as of mid-March 2021, we would have been halfway through this pandemic in terms of cumulative deaths. Moreover, absent the development of technological solutions, my model implies sharply diminishing returns to further non-pharmaceutical interventions in reducing the long-run death toll from COVID-19 even if such measures had been implemented early in 2020 and maintained for a long, but finite, length of time. Thus, absent a technological solution, we would be faced with few options for further mitigating the long-run impact of COVID-19 on public health.

Second, here in the United States, we have been very fortunate with our success in developing and now implementing effective vaccines against COVID-19. With vaccines, the long-run death toll from COVID-19 is forecast to be roughly 600,000, or about half the level without such a technological solution. This forecast takes into account both the relaxation of private and public efforts at disease control that we have seen in spring 2021 and the arrival of new, more contagious variants of the virus. Clearly, Operation Warp Speed and the associated research effort has been a scientific and public health achievement of historic importance.

Third, in contrast to the case of no technological solutions being developed, strong non-pharmaceutical interventions implemented early on are highly complementary with speedy development of vaccines and life-saving therapeutics in that they save lives by delaying illness and death until such technological solutions are available. This model forecast that plausible additional non-pharmaceutical interventions, applied early on and consistently over time on top of the policies that were implemented at state and local levels, could have reduced the long-term death toll from COVID-19 in the United States to roughly 300,000 over a five-year period. This forecast takes into account the likely countervailing relaxation of private and state and local mitigation efforts had such interventions been implemented at a federal level. Based on this forecast, I conclude that here in the United States, over the course of the past 14 months, we failed to take actions that would have saved hundreds of thousands of lives. Given the success of a number of countries in containing COVID-19 over the past year while preserving economic activity, it is entirely plausible that such

non-pharmaceutical interventions would not have led to high economic costs and, in fact, might have led to better economic outcomes.

Fourth, and finally, looking ahead, we face a future in which COVID-19 will remain a threat as long as it is prevalent elsewhere in the world and in which new pandemic threats will likely arise. We should use the worldwide experience with COVID-19 to guide investments in public health infrastructure that will allow us to rapidly identify and react to pandemic threats with more effective and less costly non-pharmaceutical interventions. Such investments have a strong "public good" rationale and would be highly complementary with increased investments in the scientific and clinical research infrastructure to rapidly develop technological solutions such as vaccines and life-saving therapeutics for future threats from infectious disease.

I. Epidemic Dynamics with and without Behavior

The public health policies enacted around the world to combat COVID-19 have been guided by standard epidemiological models built on the susceptible-infected-removed (SIR) framework developed by Kermack and McKendrick (1927). These models simulate disease transmission as arising when infectious individuals (corresponding to the *I* in SIR) interact with others. Through this interaction, a virus or other pathogen succeeds in infecting those who have no immunity and are thus susceptible (corresponding to the *S* in SIR), turning such agents into newly infectious individuals. Individuals who gain immunity from prior infections or vaccinations are said to be removed (corresponding to the *R* in SIR) as they no longer contribute to the transmission of the disease. The progress of the epidemic through the population is mechanical as the rate at which infectious people interact with others is assumed to be invariant to the current prevalence of the disease.

When applied to COVID-19, three quantitative implications of this standard model stand out.[3] First, the model gives dire forecasts for the peak of the disease's first wave—10 to 20 percent of Americans were predicted to be sick with COVID-19 simultaneously at the first peak of infections absent drastic efforts such as lockdowns to slow transmission. At current estimates of the infection fatality rate for COVID-19, this rate of infection would have corresponded to peak death rates on the order of 30,000 to

3. See Atkeson (2020) and Stock (2020) for expositions of these predictions of standard SIR models from one year ago.

60,000 deaths per day.[4] Second, this standard model forecast that if efforts to slow transmission through lockdowns were applied early on but were only temporary, this dramatic first peak would be delayed but not prevented: cases and deaths would explode again once efforts to slow transmission were relaxed. Third, this standard model offered dramatic long-run predictions of the kind made famous by Angela Merkel in March 2020—that more than two-thirds of Germany's population were forecast to experience infections (if not vaccinated) before the pandemic would end through herd immunity.[5] Again, applying current estimates of the average infection fatality rate for COVID-19 in the United States, this implies a long-run death toll on the order of 1.49 million or more.[6]

These implications of a standard epidemiological model for the magnitude of the first peak and the long-run impact of COVID-19 in terms of infections are driven by a single parameter known as the basic reproduction number of the virus (the R_0).[7] The implications of these infections for deaths from COVID-19 are determined by the average infection fatality rate across the infected population. While we now know that the infection fatality rate from COVID-19 varies widely with age and other factors, estimates of the disease burden from COVID-19 from the Centers for Disease Control and Prevention (CDC) are consistent with an average infection fatality rate of 0.5 percent across the entire infected population in the United States for 2020.[8] The emergence of new, more transmissible virus variants with higher basic reproduction numbers makes the predictions of standard epidemiological models for peak infections and long-run impact even more dire.

4. This estimate for peak deaths is likely understated given that such a wave of infections would clearly have overwhelmed the health care system.

5. See Die Bundesregierung [Cabinet of Germany], "An Address to the Nation by Federal Chancellor Merkel," n.d., https://www.bundesregierung.de/breg-de/themen/coronavirus/statement-chancellor-1732296; and CNN, "Merkel Believes 60–70% of Germany's Population Will Be Infected," March 11, 2020, https://edition.cnn.com/world/live-news/coronavirus-outbreak-03-11-20-intl-hnk/h_ab9bb8236fa91a9bf63cdbc7a69e0f10.

6. This forecast for the cumulative death toll in this model scenario does not take into account that the infection fatality rate would likely have risen substantially due to congestion in the health care system if the first wave of infections had approached anything close to the levels forecast by this standard model.

7. See Randolph and Barreiro (2020) for a description of the calculations and considerations involved.

8. The CDC estimates that 83 million Americans had been infected by the end of December 2020. Total COVID-19 deaths reached 445,000 thirty days later, giving an average estimated infection fatality rate, including the delay from infection to death, of slightly over 0.005. See CDC, "Estimated Disease Burden of COVID-19," https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/burden.html.

It is now clear that the first prediction of standard epidemiological models for the first peak of infections and deaths was off by at least an order of magnitude—it is unlikely that more than 2 percent of Americans have ever been infected simultaneously, and the peak of daily deaths in America from COVID-19 has fortunately stayed under 4,000. Looking at data worldwide, it appears that the second prediction of standard epidemiological models is also off, perhaps by an order of magnitude. While many locations within the United States and abroad have suffered severe second or third waves of COVID-19 deaths after relaxing costly public measures to control disease transmission, these waves have been much smaller than predicted by a standard SIR model.

In contrast, the standard SIR model's third prediction, regarding long-run impact, looks to be closer to the mark. While the precise threshold of herd immunity—the fraction of the population that has to gain immunity through infection or vaccination before the pandemic can end—is not yet empirically resolved, available data from locations such as Manaus, Brazil, which has experienced high rates of infection, and from Israel, the United Kingdom, and the United States, each of which have high vaccination rates with effective vaccines as of mid-March 2021, indicate that the predictions of a standard epidemiological model for the long-run impact of COVID-19 are likely correct: this pandemic will not resolve until high proportions of the population have acquired immunity either through infection or vaccination.[9]

I.A. Behavior Regulates Disease Dynamics

How does consideration of the impact of behavior on the progression of a pandemic help us understand this relationship between the predictions of a standard SIR model and observed outcomes?

Within economics, Tomas Philipson pioneered the study of the interaction of behavior and the spread of disease in his work on the HIV/AIDS pandemic. In a chapter in *Handbook of Health Economics*, summarizing work on that pandemic, Philipson (2000) argued that epidemiological models should incorporate prevalence-elastic private demand for costly measures to prevent the spread of infectious disease. Such models, he maintained, offered two fundamental economic insights.

9. See Sabino and others (2021) and Mallapaty (2021) regarding data from Manaus and Israel on the empirical herd immunity threshold. The CDC COVID-19 data tracker site reports on vaccination rates in the United States. Data on vaccination rates in the United Kingdom are available at Gov.UK, "Coronavirus (COVID-19) in the UK," https://coronavirus.data.gov.uk/details/vaccinations.

The first insight is that costly private efforts to prevent disease transmission are self-limiting—as disease incidence falls, these costly efforts to control disease spread are relaxed and the disease reemerges.

Within the United States, it appears that this observation holds for public policies aimed at COVID-19 as well—state and local disease control measures are often conditioned on measures of disease prevalence such as infections or hospitalizations, and these public measures aimed at the control of COVID-19 are relaxed as disease prevalence falls. In the model I present below, I interpret this correlation between public policies and disease prevalence as arising from a public behavioral response to shifting political calculations as disease prevalence rises and falls, that is, as a social-choice behavioral response that is conceptually similar to private behavioral responses. I thus interpret the reduced form behavioral response of transmission rates to disease prevalence in my model as resulting from a combination of private and public reactions to disease prevalence.

The second insight is that the private and public behavioral response to changing disease prevalence partially offsets the impact of additional non-pharmaceutical interventions aimed at disease control. In short, the effect of a specific non-pharmaceutical intervention is limited by its success as private and public efforts aimed at disease control are relaxed in response.

That both public and private prevalence-elastic demand for costly measures to control disease are self-limiting is a particularly powerful insight for understanding where the standard epidemiological model fails as a description of disease dynamics and where it succeeds. In joint work with Karen Kopecky and Tao Zha (Atkeson, Kopecky, and Zha 2021), we find that the data on the progression of the COVID-19 pandemic across many countries and US states throughout 2020 conform strikingly well to a core prediction of the standard epidemiological model modified to include prevalence-elastic demand for disease prevention—that after the first phase of the pandemic in which disease grows rapidly, the growth rates of infections and deaths should remain in a relatively narrow band around zero until the pandemic is over.[10]

The intuition for this prediction regarding disease dynamics in the context of a model with prevalence-elastic demand for disease prevention is simple. If new infections and daily deaths from the disease grow too high, people and governments make costly efforts to avoid interaction and thus slow

10. Joshua Gans (2020) reviews the implications of epidemiological models with a prevalence-elastic demand for costly measures to prevent disease transmission and much of the work by others on this topic.

disease spread. Likewise, if the prevalence of the disease falls, people and governments relax those costly measures to prevent disease transmission and the prevalence of the disease rises again. The reaction of behavior, both public and private, to the prevalence of the disease regulates the equilibrium prevalence of the disease in the same way that cruise control regulates the velocity of a car on a highway that winds up and down hills. The equilibrium level of daily deaths, corresponding in this analogy to the velocity of the car, remains within a relatively narrow band (relative to that predicted by a standard SIR model) in response to shocks having an impact on disease transmission because of the stabilizing role of endogenous prevalence-elastic public and private disease avoidance behavior. The impact of this behavior then is to transform what would otherwise be a short and sharp disease episode into a much more slowly evolving and drawn-out phenomenon.

What are the implications of a model with prevalence-elastic demand for disease prevention for the long-run impact of an epidemic? Here the insight that the demand for disease prevention is self-limiting is particularly relevant. For an epidemic to end, the prevalence of the disease must fall toward zero. As disease prevalence falls toward zero, the demand for costly disease prevention efforts also falls toward zero, and hence the disease will come back unless the population has already achieved herd immunity measured at prepandemic levels of behavior. That is, the predictions for the long-run impact of COVID-19 using a standard epidemiological model should continue to hold.[11] Given estimates of the basic reproduction number around 3, or now higher with new variants, this herd immunity threshold should kick in when significantly less than one-third of the population remains susceptible.[12]

This logic implies that, absent a vaccine or the development of life-saving therapeutics, the implications of a model that includes a prevalence-elastic demand for disease prevention for the long-run impact of a pandemic in terms of cumulative infections and deaths should be similar

11. More complex models that emphasize heterogeneity and the network structure of human interaction potentially offer more optimistic implications for the long-run impact of COVID-19. See, for example, Ellison (2020), Akbarpour and others (2020), Azzimonti and others (2020), and Boppart and others (2020).

12. On the transmissibility of the UK variant, see Davies and others (2021); for the even higher transmissibility of the Indian variant, see "Delta Coronavirus Variant Believed to Have 60% Transmission Advantage—UK Epidemiologist," Reuters, June 9, 2021, https://www.reuters.com/business/healthcare-pharmaceuticals/delta-coronavirus-variant-believed-have-60-transmission-advantage-uk-2021-06-09/.

to that of a standard epidemiological model. The slowing of the epidemic that results from a behavioral response to disease prevalence can reduce the cumulative death toll by reducing the extent to which cumulative infections in the long run overshoot the herd immunity threshold, but this behavioral response does not reduce the cumulative impact of the epidemic to a point below this threshold. In the case of COVID-19 in the United States, in the model I present below, this would be a cumulative death toll on the order of 1.24 million.

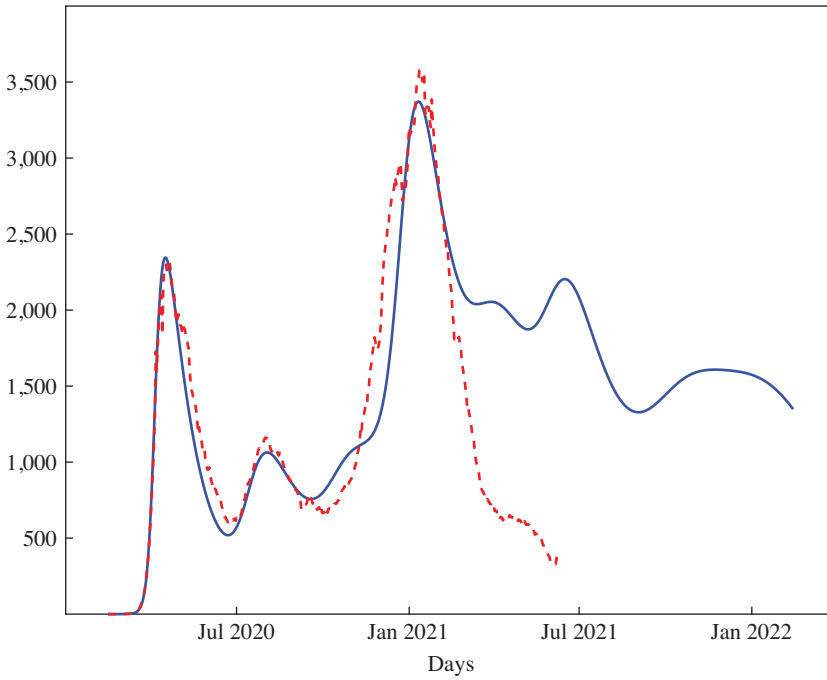
I.B. A Quantitative Illustration

To illustrate these points regarding the predictions of a standard epidemiological model and one with a prevalence-elastic demand for disease prevention for the dynamics of an epidemic, I turn to a model of the dynamics of deaths from the COVID-19 epidemic in the United States that I presented in a recent working paper (Atkeson 2021) and which is included as an online appendix to this paper. This model accounts for the dynamics of deaths from COVID-19 in the United States over the past year with shocks to transmission rates due to seasonality, due to the emergence of a new, more transmissible variant of the novel coronavirus, and due to potential changes in the prevalence-elasticity of demand for costly measures to mitigate disease transmission. (I refer to this third shock as "pandemic fatigue" as a shorthand description of a decline in the responsiveness of private and public demand for costly disease prevention measures to changes in disease prevalence. This shock is perhaps a reduced form for a more dynamic response of behavior as a pandemic wears on.)

This model accounts remarkably well for the pandemic's evolution in the United States over the past year. In the online appendix, I document that, in the model, a seasonal decline in transmission rates explains why the prevalence of COVID-19 dropped to relatively low levels in the summer of 2020. In the model, a decline in the strength of the behavioral response to disease prevalence in late fall—pandemic fatigue—explains the large waves of infections and deaths seen in the late fall and winter. The introduction of a more transmissible variant in early December together with the start of an aggressive vaccination program explain the progress of the epidemic in the spring of 2021.[13]

13. In the online appendix, I document the specific features of this model that allow it to fit the pattern of daily deaths observed over the past 14 months with relatively few shocks and discuss the procedure used to choose the model parameters. The fit of the model to the data is serendipitous. Further research is needed to develop behavioral models that can fit the wide range of experiences with COVID-19 seen across regions and countries of the world.

Figure 1. Daily Deaths with a Behavioral Response but No Vaccines



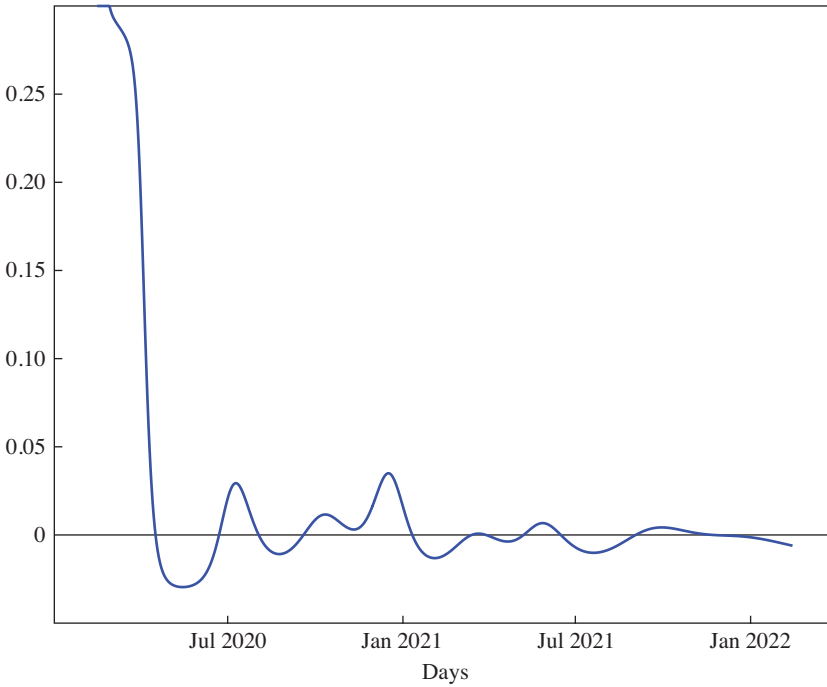
Sources: CDC and author's calculations.

Notes: Behavioral model implications for daily deaths in the United States from mid-February 2020 through mid-February 2022 are shown in the solid line. Seasonal variation, the introduction of a more contagious variant in December 2020, and prevalence-elastic demand for costly measures to slow disease transmission have an impact on transmission rates. The onset of pandemic fatigue in late 2020 accounts in large part for the peak in deaths in January 2021. Data on the seven-day moving average of daily deaths are shown in the dashed line. The forecast for cumulative deaths over a five-year period implied by this model is 1.24 million.

In figure 1, I show the model's prediction for daily deaths from COVID-19 in the United States from mid-February 2020 to mid-February 2022 (the solid line), and data on the seven-day moving average of daily deaths in the United States over the past year (the dashed line) downloaded from the CDC's COVID-19 data tracker website.[14] The behavioral model matches the data on deaths over the past year quite well, and it forecasts,

14. CDC, "COVID Data Tracker," https://covid.cdc.gov/covid-data-tracker/#datatracker-home. Note that these data on daily deaths omit roughly 14,000 deaths included in the CDC estimate of the cumulative death toll from COVID-19 available on the same site as these additional deaths were included retroactively due to reclassification of state and local death counts.

Figure 2. Growth Rate of Daily Deaths



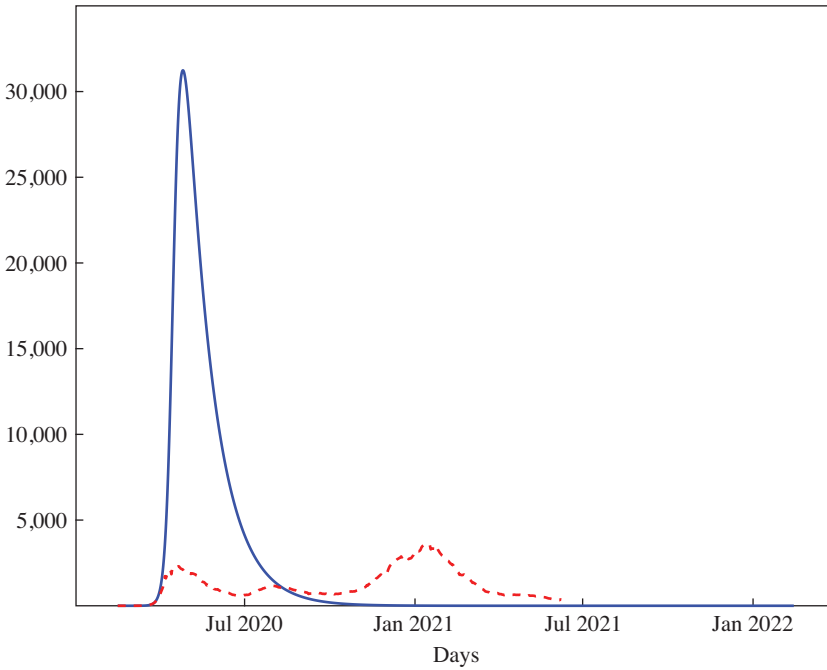
Source: Author's calculations.

Notes: The growth rate of daily deaths implied by the solution of the model shown in figure 1 shows that behavior closely regulates the growth rate of the epidemic after its initial phase of rapid growth.

absent vaccines, a continuation of the pandemic well into 2022. The predicted path of daily deaths through 2021 shown in this figure is driven by the spread of the new, more contagious virus variant in the model. The long-run cumulative death toll over a five-year period in this forecast run of the model in figure 1 is 1.24 million. The forecast shown in this figure does not include any consideration of the impact of vaccines, both to permit comparison with projections from a standard epidemiological model and to serve as a benchmark for the impact of vaccination efforts.

To illustrate the impact of the behavioral response to disease prevalence in shaping the growth rate of the epidemic, in figure 2 I show the model-implied growth rate of daily deaths from the simulation of the model shown in figure 1. We see in this figure that the growth rate of daily deaths starts out at a very high level—above 30 percent per day—and then falls rapidly toward zero and hovers around zero even with shocks due to seasonality

Figure 3. Daily Deaths with No Behavioral Response or Vaccines



Sources: CDC and author's calculations.

Notes: Standard model implications for daily deaths in the United States from mid-February 2020 through mid-February 2022 are shown in the solid line. Seasonal variation and the introduction of a more contagious variant in December 2020 have an impact on transmission rates, but this specification of the model has no prevalence-elastic demand for costly measures to slow disease transmission. Data on the seven-day moving average of daily deaths in the United States over the past year are shown in the dashed line. The forecast for cumulative deaths over this five-year period is 1.49 million.

in transmission, pandemic fatigue, and the introduction of new variants. In the model, the response of private and public behavior to the level of daily deaths acts to slam the brakes on the growth of the epidemic in its initial phase and then maintain that growth rate of daily deaths in a narrow band around zero in the face of shocks to transmission much as cruise control regulates the acceleration of a car on the highway.

To contrast the implications of this model incorporating a behavioral response to disease prevalence with the implications of a standard model without such a response, in figure 3, I show the prediction for daily deaths of the same model with the behavioral response to disease prevalence turned off (the solid line), relative to data on the seven-day moving average of daily deaths (the dashed line). As we see in this figure, the standard

epidemiological model without a behavioral response overstates the first peak of daily deaths by at least an order of magnitude (these peak at over 30,000 per day), but then the pandemic comes quickly to an end in the fall of 2020. The cumulative death toll in this model forecast is 1.49 million. This prediction for the cumulative death toll is certainly larger than in the model with a behavioral response, but the gap between the two models in this dimension is much smaller than in their predictions for the initial peak and the time scale of the pandemic.[15]

What is evident from these figures is that incorporating a response of public and private behavior to disease prevalence gives a dramatically different forecast for the severity of disease peaks as well as for the speed with which this epidemic passes through the population. In this behavioral model, absent the introduction of vaccines, the pandemic takes two and a half years to play out rather than six to nine months as forecast by the standard model without consideration of behavior. The model's implications, however, for the long-run impact of the disease are not much altered by the consideration of behavior. In both basic and behavior variations, the model forecasts that a substantial majority of the population must become immune through infection or vaccination for the pandemic to end.

II. Private Behavior and Constraints on Policy

Given these insights on the impact of prevalence-elastic demand for disease prevention on the dynamics of an epidemic, what are our options for using public policy to mitigate the impact of a pandemic on public health? One insight that I have already mentioned is that there is likely to be an offsetting private behavioral response to public measures that limit the spread of disease—that is, that additional non-pharmaceutical interventions to control an epidemic may well be partially undone by private responses and the responses of other government actors to declining disease prevalence. The other insight is that public measures of disease prevention must be essentially permanent to result in a meaningful reduction of the long-run impact of an epidemic absent a technological solution such as a vaccine or the development of life-saving therapeutics.

We can use our simple behavioral model to illustrate the quantitative implications of these two insights. Imagine that through public policies

15. This difference between the cumulative death toll forecast in the model run in figure 3 and that in figure 1 is due to what is known as "overshooting" of herd immunity in the model without behavior in figure 3. See Bergstrom and Dean (2020) for an explanation of this concept.

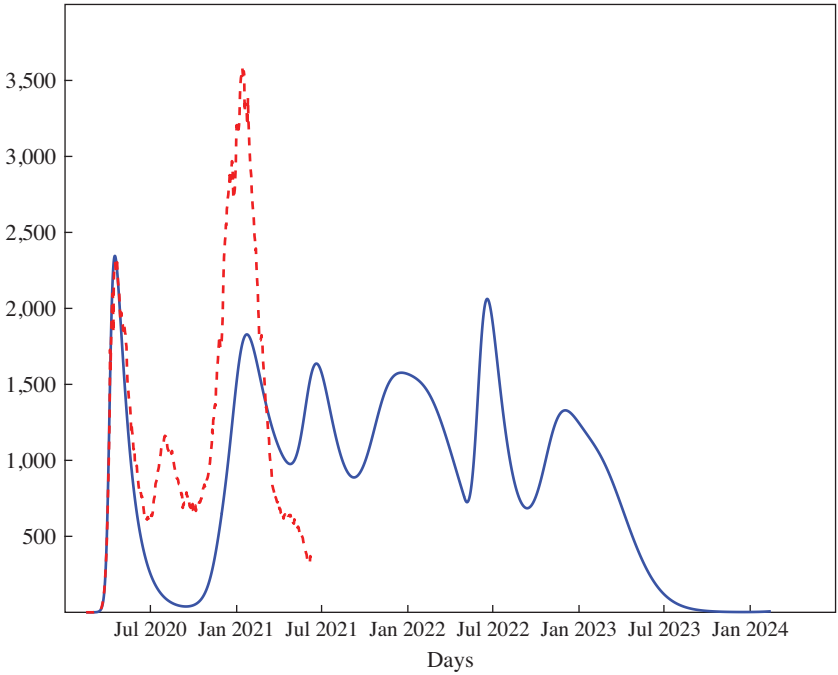
facilitating a wide range of disease control measures such as mask wearing and social distancing protocols, testing and contact tracing with isolation of infectious persons, and other measures, it was possible to significantly reduce the transmission rate of COVID-19, holding fixed seasonality and the level of costly disease control measures undertaken by both private agents and state and local authorities. Imagine that these policy interventions are undertaken for a fixed period of time independent of disease prevalence. In this sense, I imagine that these interventions are undertaken independently of the political process that leads currently observed public interventions to rise and fall with disease prevalence. Here, for purposes of illustration, I imagine these interventions as being carried out by the federal government.

In figure 4, I show a simulation of the model with such measures put in place for a two-year period from May 1, 2020, through May 1, 2022. Here I assume that these additional mitigation measures are put in place independent of the level of daily deaths and that they act to reduce disease transmission by 40 percent—a factor of $exp(-0.5)$—on top of whatever reductions in transmission are brought about by private and public changes in behavior undertaken in response to disease prevalence. I show the model implications for daily deaths over a four-year period as a solid line and the data on the seven-day moving average of daily deaths in a dashed line. As we see in this figure, these disease control measures, when imposed on top of those arising in equilibrium from the prevalence-elastic demand of both private agents and public authorities for costly measures to control disease, have a significant impact in reducing deaths from the disease in the first year. Then, in this simulation, in early 2021, the arrival of the new variant and, in mid 2022, the abandonment of these disease control measures leads to significant spikes in forecast deaths. Over the long run, the cumulative death toll is 1.22 million—almost exactly what we found in the simulation in figure 1 that had no such additional disease control measures imposed. This simulation indicates sharply diminishing returns to additional nonpharmaceutical interventions absent a technological solution such as a vaccine or life-saving therapeutics.

II.A. Waiting for a Technological Solution

We saw in figure 4 that additional but temporary disease control measures do not significantly reduce the long-run public health impact of the epidemic in the absence of a technological solution such as vaccines or life-saving therapeutics. How does the analysis of the impact of such measures change when there is a good prospect that a vaccine or therapeutics

Figure 4. Daily Deaths with Extra Mitigation but No Vaccines



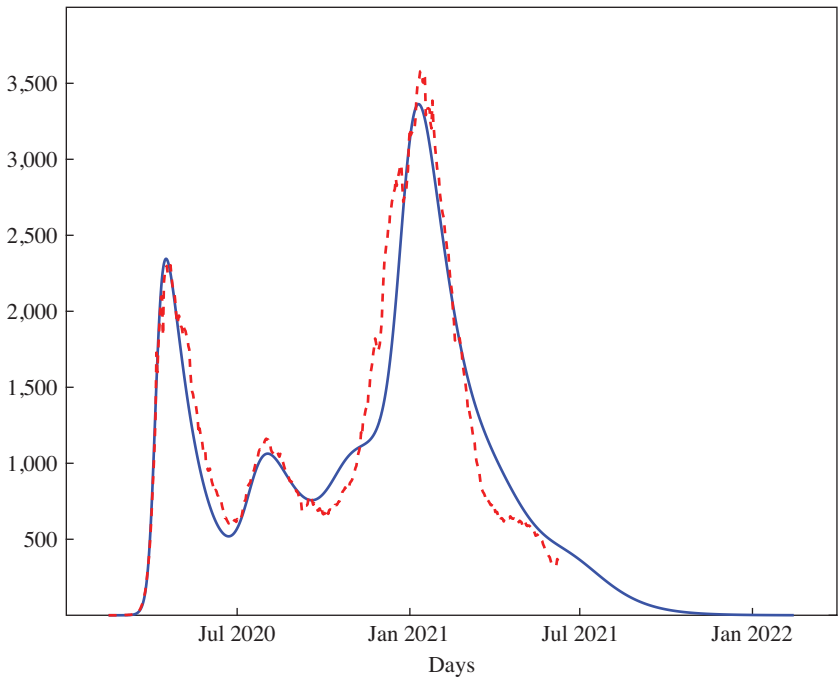
Sources: CDC and author's calculations.

Notes: Predictions of the model for the evolution of daily deaths from COVID-19 in a version in which disease control measures—such as masks, social distancing, testing with contact tracing, and isolation of the infected—cut the transmission rate of the disease by 40 percent, a factor of $exp(-0.5)$, holding fixed the level of private and state and local disease control efforts undertaken in response to the prevalence of the disease. These measures are assumed to be in place for two years from May 1, 2020, to May 1, 2022. While these disease control measures are effective in reducing deaths in the first year, they do not succeed in later years. The data on the seven-day moving average of daily deaths are shown in the dashed line. The cumulative death toll implied by this model is 1.22 million—nearly the same as shown in the simulation in figure 1.

might arrive? Here I use the model to show that such measures can have a significant long-run public health benefit in reducing deaths from disease while waiting for the arrival of that technological solution.

In figure 5, I show the implications of the model for the evolution of daily deaths (the solid line) when a program of vaccination starts on January 1, 2021, at a pace sufficient to succeed in protecting half of the US population by July 1, 2021. This vaccine is assumed to prevent both illness and disease transmission by the vaccinated. The data on the seven-day moving average of daily deaths are again shown in a dashed line. To

Figure 5. Daily Deaths with a Behavioral Response and Vaccines



Sources: CDC and author's calculations.

Notes: Predictions of the model for daily deaths from COVID-19 (in the solid line) in a simulation with a vaccination program starting on January 1, 2021, that proceeds at a rate fast enough to protect half of the population by July 1, 2021. In this simulation, the vaccine is assumed to protect against illness and to prevent disease transmission by the vaccinated. The data on the seven-day moving average of daily deaths are shown in the dashed line. To see the predicted impact of the vaccine on the dynamics of the epidemic, compare the solid line for model-implied daily deaths in 2021 in figure 1 to the solid line here. The cumulative death toll over a five-year period in this simulation is 595,000.

see the model-implied impact of this vaccination program on the epidemic, one can compare the solid lines in figures 1 and 5. Here we see that, in the model, this vaccination program significantly reduces the forecast impact of the new variant in late spring 2021 and brings the epidemic to an end late by summer or fall 2021. Note that here the vaccination program succeeds despite the model-implied relaxation of public and private efforts at disease prevention in response to falling daily deaths.

The long-run death toll predicted by the model with this vaccination program is 595,000, less than half of what is forecast in the absence of a vaccine (in the simulations in figures 1 and 4). In this sense, the vaccination

program succeeds in substantially reducing cumulative deaths in a manner that a two-year program of disease mitigation absent a vaccine does not.

But now consider the model-implied scenario for cumulative deaths if the temporary disease mitigation measures used in the simulation in figure 4 had been imposed starting May 1, 2020, and the same vaccination program applied in the simulation in figure 5 had started on January 1, 2021. With this combination of temporary disease mitigation measures and a successful vaccination program, the cumulative death toll implied by the model would have been only 302,000. Clearly, the combination of temporary disease control measures applied while waiting for a technological solution can save many lives. The lesson here is that there are tremendous complementarities between early and aggressive mitigation and the development of a technological solution such as vaccines or life-saving therapeutics in terms of reducing the public health impact of a pandemic.

III. Conclusion

The global COVID-19 pandemic has clearly demonstrated that the risks from the emergence of new infectious diseases, which epidemiologists have been speaking about for years, are terribly real. This pandemic has also posed a severe test of public health strategies and capabilities worldwide. In many countries, the associated economic impact has been as severe as any downturn seen since the Great Depression. How might we do better next time?

Based on the lessons about the interaction of behavior and disease dynamics discussed here, I suggest the following three-part strategy to improve our public health and economic response to emerging infectious disease.

First, we need to invest in our disease surveillance capabilities worldwide, perhaps using the infrastructure developed for worldwide influenza surveillance as a model.[16] It is certainly worth a lot of money to have the capacity to identify the threat from a new infectious disease anywhere in the world before it gets going so as to buy time to mount a public health and scientific response.

Second, we must consolidate all that has been learned about the implementation of non-pharmaceutical public health measures for disease control

16. World Health Organization, "Preventing the Next Human Influenza Pandemic: Celebrating 10 Years of the Pandemic Influenza Preparedness Framework," https://www.who.int/news/item/21-05-2021-preventing-the-next-human-influenza-pandemic-celebrating-10-years-of-the-pandemic-influenza-preparedness-framework.

over the past year so that we might be able to quickly implement those measures that have been proven to effectively slow the spread of an infectious disease with the least cost to the economy. Given the widespread discussion of pandemic fatigue in the population, we should also look at policies for infectious disease control that have low personal costs and thus have a greater chance of enjoying widespread voluntary adherence. A number of countries, many of them in Asia, have been able to keep COVID-19 infections and deaths to low levels over the course of the past year with effective public health interventions based on travel restrictions, testing, contact tracing, and isolation of infected individuals while preserving considerable economic activity and personal autonomy. Several universities in the United States have also succeeded at control of COVID-19 infections with extensive testing and isolation regimes.[17] As we have seen from the model simulations in this paper and these real-world experiences, public health measures that allow us to wait for the development of a technological solution to a global pandemic with minimal loss of life and economic damage can be extremely valuable. Given the public-good nature of infectious disease surveillance and public health system preparedness to implement rapidly scalable countermeasures, it seems a high priority to fund these capabilities at the federal level.

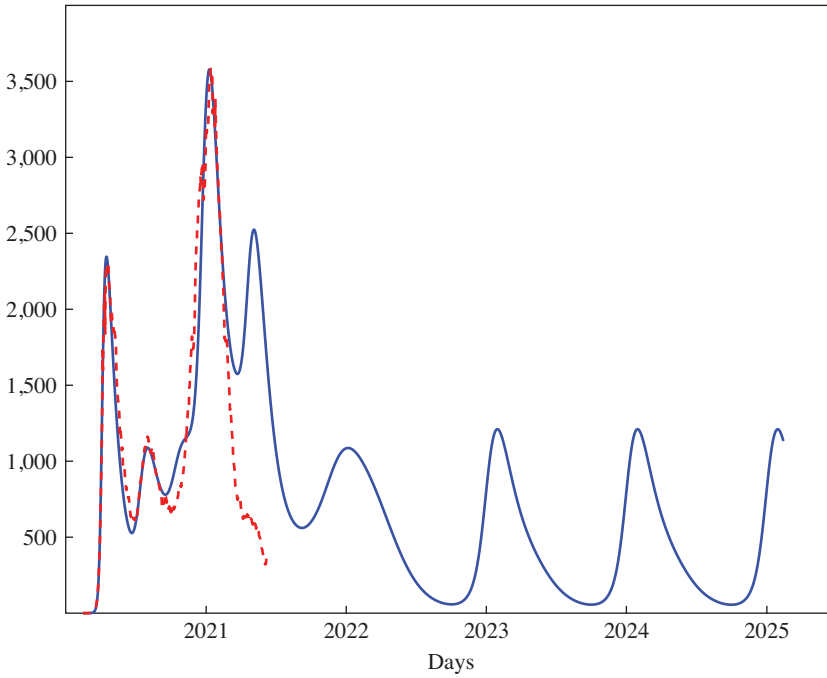
Third, we need to invest in new models for accelerating the development, financing, and distribution of vaccines and life-saving therapeutics for emergent disease. In the end, it is these technological solutions that will allow us to contain the long-run impact of new pandemics once they become global.[18]

To illustrate the urgency of addressing these public health priorities now, consider one final model scenario. As long as COVID-19 remains prevalent worldwide, new mutations of the virus are likely to emerge and there is increasing evidence that such mutations might allow COVID-19 to evade the immunity conferred by prior infection and vaccines. In such a scenario, COVID-19 could be an endemic seasonal disease that might

17. See, for example, the experience of the University of Illinois at Urbana-Champaign and that of Georgia Tech: University of Illinois, "On-Campus COVID-19 Testing Data Dashboard," https://covid19.illinois.edu/on-campus-covid-19-testing-data-dashboard/; Georgia Tech, "Covid-19 Exposure and Health Alerts," https://health.gatech.edu/coronavirus/health-alerts.

18. See Council of Economic Advisers (2019) for a careful analysis of the economic and public health rationale for a large federal investment in such technologies. See Angus, Gordon, and Bauchner (2021) for a discussion of current difficulties in conducting rapid clinical trials of new treatments in the United States.

Figure 6. Daily Deaths with Waning Immunity



Source: Author's calculations.

Notes: Predictions of the model for daily deaths from COVID-19 (solid line) in a simulation with a vaccination program starting on January 1, 2021, that proceeds at a rate fast enough to protect half of the population by July 1, 2021. The data on the seven-day moving average of daily deaths are shown in the dashed line. In this simulation, immunity acquired from prior infection or vaccination is assumed to last 18 months on average. The vaccination program is assumed to continue at a constant rate throughout the entire period with new booster shots conferring immunity against new variants as they occur. Even with a program of booster vaccines and continued prevalence-elastic behavior, in this simulation roughly 100,000 Americans die each year from COVID-19 on a persistent basis.

require essentially permanent efforts at disease control.[19] To illustrate how such a scenario might play out, I simulate the model with vaccines shown in figure 5 with a version of the virus circulating that is two-thirds more transmissible than the original virus, but in which immunity from infection or vaccination lasts on average for only 18 months. I show the resulting forecast path of daily deaths in the United States from COVID-19 over a five-year period in figure 6. In this simulation, I assume that the vaccination

19. See, for example, Murray and Piot (2021); see also Lavine, Bjornstad, and Antia (2021).

program continues at a constant rate of roughly 1.3 million vaccinations per day throughout the forecast period. As one can see in this figure, the epidemic is forecast in this scenario to settle into a regular seasonal pattern killing over 100,000 Americans per year even with new vaccines and a response of public and private behavior to the changing prevalence of the disease. Clearly, in such a scenario, we would benefit greatly from finding ways to mitigate this disease on an ongoing basis at a lower economic cost.

ACKNOWLEDGMENTS I would like to thank Michael Droste, Karen Kopecky, Ben Moll, Steven Shafer, James Stock, and Tao Zha for many useful conversations.

References

Akbarpour, Mohammad, Cody Cook, Aude Marzuoli, Simon Mongey, Abhishek Nagaraj, Matteo Saccarola, Pietro Tebaldi, Shoshana Vasserman, and Hanbin Yang. 2020. "Socioeconomic Network Heterogeneity and Pandemic Policy Response." Working Paper 27374. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27374.

Angus, Derek C., Anthony C. Gordon, and Howard Bauchner. 2021. "Emerging Lessons from COVID-19 for the US Clinical Research Enterprise." *JAMA* 325, no. 12: 1159–61.

Armstrong, Gregory L., Laura A. Conn, and Robert W. Pinner. 1999. "Trends in Infectious Disease Mortality in the United States during the 20th Century." *JAMA* 281, no. 1: 61–66. https://doi.org/doi:10.1001/jama.281.1.61.

Atkeson, Andrew. 2020. "What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios." Working Paper 26867. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26867.

Atkeson, Andrew. 2021. "A Parsimonious Behavioral SEIR Model of the 2020 COVID Epidemic in the United States and the United Kingdom." Working Paper 28434. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28434.

Atkeson, Andrew, Karen A. Kopecky, and Tao Zha. 2021. *Behavior and the Transmission of COVID-19*. Staff Report 618. Minneapolis: Federal Reserve Bank of Minneapolis. https://www.minneapolisfed.org/research/staff-reports/behavior-and-the-transmission-of-covid19.

Azzimonti, Marina, Alessandra Fogli, Fabrizio Perri, and Mark Ponder. 2020. "Pandemic Control in ECON-EPI Networks." Working Paper 27741. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27741.

Bergstrom, Carl T., and Natalie Dean. 2020. "What the Proponents of 'Natural' Herd Immunity Don't Say." *New York Times*, May 1. https://www.nytimes.com/2020/05/01/opinion/sunday/coronavirus-herd-immunity.html.

Boppart, Timo, Karl Harmenberg, John Hassler, Per Krusell, and Jonna Olsson. 2020. "Integrated Epi-Econ Assessment." Working Paper 28282. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28282.

Council of Economic Advisers. 2019. *Mitigating the Impact of Pandemic Influenza through Vaccine Innovation*. Washington: Author. https://trumpwhitehouse.archives.gov/wp-content/uploads/2019/09/Mitigating-the-Impact-of-Pandemic-Influenza-through-Vaccine-Innovation.pdf.

Davies, Nicholas G., Sam Abbott, Rosanna C. Barnard, Christopher I. Jarvis, Adam J. Kucharski, James D. Munday, and others. 2021. "Estimated Transmissibility and Impact of SARS-CoV-2 Lineage B.1.1.7 in England." *Science* 372, no. 6538. https://science.sciencemag.org/content/372/6538/eabg3055.

Ellison, Glenn. 2020. "Implications of Heterogeneous SIR Models for Analyses of COVID-19." Working Paper 27373. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27373.

Gans, Joshua S. 2020. "The Economic Consequences of $\hat{R} = 1$: Towards a Workable Behavioral Epidemiological Model of Pandemics." Working Paper 27632. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27632.

Kermack, William Ogilvy, and A. G. McKendrick. 1927. "A Contribution to the Mathematical Theory of Epidemics." *Proceedings of the Royal Society of London A* 115: 700–721.

Lavine, Jennie S., Ottar N. Bjornstad, and Rustom Antia. 2021. "Immunological Characteristics Govern the Transition of COVID-19 to Endemicity." *Science* 371, no. 6530 (February): 741–45. https://science.sciencemag.org/content/371/6530/741.

Mallapaty, Smriti. 2021. "Vaccines Are Curbing COVID: Data from Israel Show Drop in Infections." *Nature*, February 5. https://www.nature.com/articles/d41586-021-00316-4.

Morens, David M., and Anthony S. Fauci. 2020. "Emerging Pandemic Diseases: How We Got to COVID-19." *Cell* 182, no. 5: 1077–92. https://doi.org/10.1016/j.cell.2020.08.021.

Murray, Christopher J. L., and Peter Piot. 2021. "The Potential Future of the COVID-19 Pandemic: Will SARS-CoV-2 Become a Recurrent Seasonal Infection?" *JAMA* 325, no. 13: 1249–50.

Philipson, Tomas. 2000. "Economic Epidemiology and Infectious Diseases." In *Handbook of Health Economics, Volume 1B*, edited by Anthony J. Culyer and Joseph P. Newhouse. Amsterdam: North-Holland.

Randolph, Haley E., and Luis B. Barreiro. 2020. "Herd Immunity: Understanding COVID-19." *Primer* 52, no. 5: 737–41.

Sabino, Ester C., Lewis F. Buss, Maria P. S. Carvalho, Carlos A. Prete Jr., Myuki A. E. Crispim, Nelson A. Fraiji, and others. 2021. "Resurgence of COVID-19 in Manaus, Brazil, Despite High Seroprevalence." *Lancet* 397, no. 10273: 452–55.

Saelens, Xavier, and Bert Schepens. 2021. "Single-Domain Antibodies Make a Difference." *Science* 371, no. 6530: 681–82.

Stock, James H. 2020. "Data Gaps and the Policy Response to the Novel Coronavirus." Working Paper 26902. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w26902.

CHRISTINA D. ROMER
University of California, Berkeley

The Fiscal Policy Response to the Pandemic

ABSTRACT This paper provides estimates of the size and determinants of the fiscal policy response to the COVID-19 pandemic across thirty advanced economies. In contrast to the fiscal response to financial crises, I find no evidence that fiscal space was an important determinant of the aggressiveness of pandemic fiscal packages. Focusing in on the US fiscal policy response, I discuss the policy implications of the unique features of a pandemic recession. I argue that the social insurance and public health components of the $5.2 trillion US package, such as expanded unemployment insurance and government funding of vaccine development and distribution, were highly appropriate, whereas broad-based stimulus measures, such as the onetime payments to households, were not. Finally, I consider some of the longer-run consequences of the US fiscal policy actions. The aggressive fiscal expansion, along with extensive private saving during the pandemic, is likely to generate rapid growth over the next few years. The rise in the debt-to-GDP ratio, caused by both the policy response and the pandemic recession itself, could limit future fiscal action if anti-debt sentiment reemerges.

The fiscal policy response to the pandemic in the United States has been extraordinary. Including the recently passed American Rescue Plan Act, pandemic-related legislation has had a budgetary cost of more than $5 trillion.[1] As a share of GDP, that is nearly equivalent to what the

Conflict of Interest Disclosure: The author did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this paper. She is currently not an officer, director, or board member of any organization with an interest in this paper.

1. Committee for a Responsible Federal Budget (CRFB), "COVID Money Tracker," https://www.covidmoneytracker.org/. The size estimate is corroborated by Gravelle and Marples (2021) and the CBO (2020, 2021).

United States spent on war production in 1943.[2] Or, to put it in a more modern context, it is about four times as large as the 2009 American Recovery and Reinvestment Act passed to help the US economy recover from the global financial crisis. Though the United States has had one of the most aggressive fiscal responses, other countries have done a great deal as well.

As we begin the second year of the pandemic, it is useful to take a step back and assess these extraordinary actions. What determined the aggressiveness of the fiscal policy response across countries? Was the composition of the US fiscal package appropriate for the special circumstances of the pandemic economy? And finally, will the fiscal response have repercussions for the future?

I. What Determined the Aggressiveness of the Fiscal Response?

I.A. Size of Early Fiscal Packages

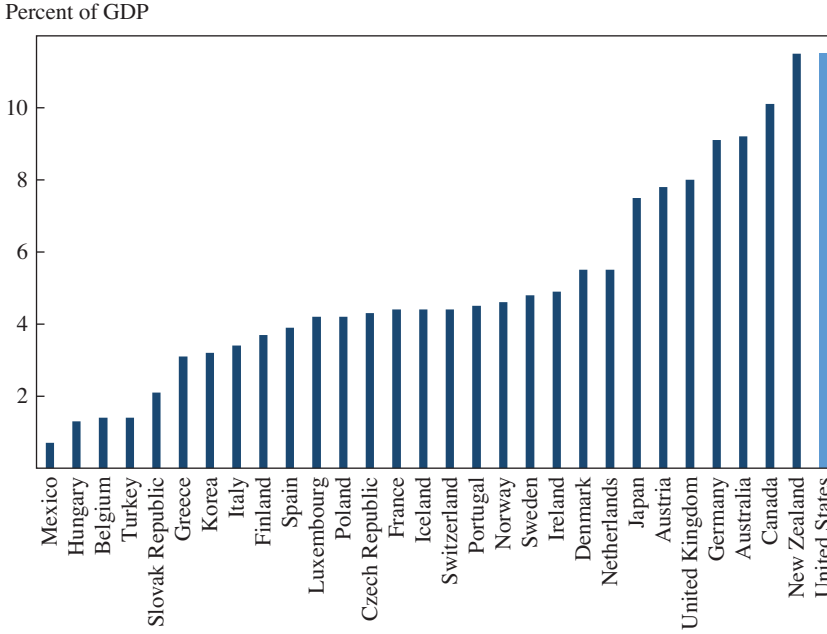
A natural place to begin is with data on the size of the fiscal policy response in various countries. David Romer and I have constructed estimates of the sizes of initial fiscal responses to the pandemic for the thirty countries in the OECD as of 2000. We aim to include only the actual budgetary impact of actions, not the headline amount of loan guarantees, liquidity provision, and similar programs. As discussed in the online appendix, we use a variety of previous fiscal policy data collection efforts (Bruegel, the IMF, and the OECD), secondary sources (Fitch Ratings and the Economist Intelligence Unit), and primary sources (country budget proposals, government announcements, and official reports) to derive our estimates of the size of fiscal packages through the end of July 2020. The online appendix describes our final adjudication for the thirty countries in our sample.

Figure 1 shows the fiscal packages (as a share of the country's 2019 GDP) ordered from lowest to highest.[3] One thing that stands out is just how extraordinary the early US fiscal response to the pandemic was. Only New Zealand spent as much relative to the size of its economy. The United States spent about 50 percent more than the United Kingdom, and roughly three times as much as France, Italy, or Spain.

2. The data on war production (based on federal contract data) were provided by Gillian Brunet.

3. We use the convention that a positive value corresponds to an increase in the budget deficit, so a larger value implies more fiscal expansion.

Figure 1. Early Pandemic Fiscal Packages in OECD Countries



Source: Author's calculations. See the online appendix for details.

The fiscal packages enacted early in the pandemic are systematically larger than early packages enacted in response to the 2008 financial crisis. The OECD (2009, 110) collected data on crisis fiscal packages in March 2009, which are similar in timing for that episode to the data we have collected for the pandemic. The average fiscal package early in the pandemic was 5.2 percent of GDP (with a median of 4.4 percent); the average package early in the Great Recession was 1.4 percent of GDP (with a median of 1.6 percent).[4] Thus, the typical package was three to four times larger in the recent episode.

I.B. Influence of Debt Ratios

In previous work, Romer and I (2018) analyzed why some countries undertook much more aggressive fiscal responses to financial crises than

4. The OECD does not include fiscal package estimates for the Great Recession episode for two countries included in our pandemic sample (Greece and Turkey). The mean pandemic package excluding Greece and Turkey is 5.4 percent (with a median of 4.5 percent).

others and, as a result, experienced much less severe post-crisis recessions. We found that a country's prior debt-to-GDP ratio had a large contractionary effect on the fiscal response to a crisis. Among OECD countries in the period since 1980, countries with initial debt ratios one standard deviation below the sample mean increased their high-employment budget deficits by over 3 percent of GDP in response to significant financial distress. On the other hand, countries with initial debt ratios one standard deviation above the sample mean actually *decreased* their high-employment deficits by 2–3 percent of GDP—meaning that they switched to highly contractionary fiscal policy.

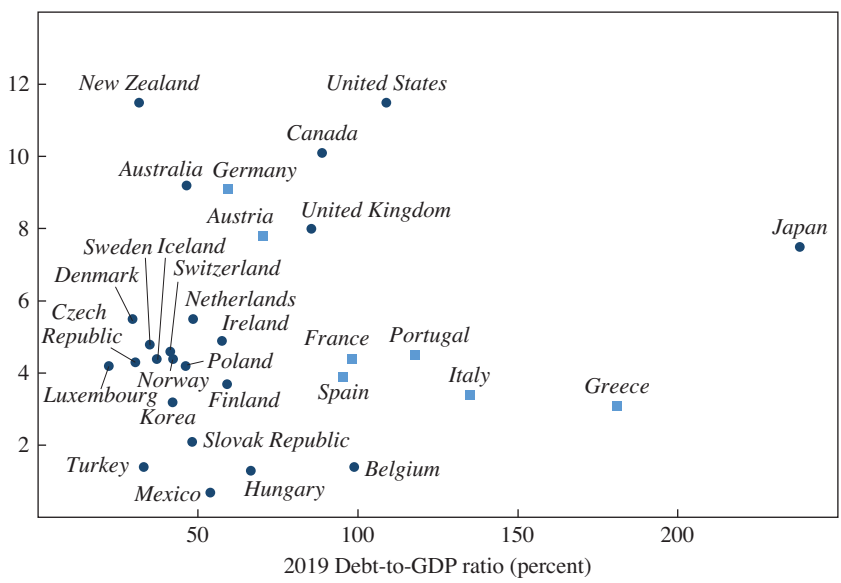
Subsequent investigation into why the fiscal response to a crisis depended on the debt ratio found only modest evidence that debt mattered because of its impact on market access (Romer and Romer 2019). For example, controlling for a country's sovereign bond rating or relative interest rate on government bonds did not noticeably reduce the impact of the debt ratio on the fiscal response to a crisis. Instead, narrative evidence suggests that "anti-debt" ideas played a crucial role. Policymakers were influenced in how they responded to a crisis by their ideas about the harms of high debt and the benefits of fiscal austerity.

Here, I examine the early COVID-19 relief packages for the same sample of countries to see if the size of the relief packages was similarly dependent on the prior debt ratio. Figure 2 shows a scatterplot of early pandemic-related fiscal packages and countries' debt-to-GDP ratios at the end of 2019.[5] There is no clear relationship between the COVID-19 relief packages and the prior debt-to-GDP ratio. Some countries with low debt, like New Zealand and Australia, took very aggressive action, but other low-debt countries, like Luxembourg and South Korea, did relatively little. At the other end of the spectrum, some high-debt countries, like Japan and the United States, did a great deal of fiscal expansion, while other high-debt countries, like Greece and Italy, did relatively little.

If one focuses on some of the core countries of the eurozone (shown in squares in figure 2), something like the expected negative relationship between debt and fiscal actions appears to hold. Low-debt Germany and Austria had early fiscal packages of about 8 percent of GDP; medium-debt France, Spain, and Portugal had packages of about 4 percent of GDP;

5. The debt data are from the International Monetary Fund (IMF) World Economic Outlook Database, October 2020. For the baseline analysis I use the gross debt ratio; I also consider the net debt ratio as a robustness exercise.

Figure 2. Scatterplot of Early Pandemic Fiscal Packages and Debt-to-GDP Ratios



Sources: Author's calculations; IMF World Economic Outlook Database.

Notes: See the online appendix for details about the fiscal response measure. The gross debt ratio data are from the IMF World Economic Outlook Database, October 2020. The countries marked with squares are the seven countries of the eurozone with the largest GDP.

and high-debt Greece and Italy had fiscal expansions of about 3 percent. However, a number of low-debt European countries, particularly the Nordic countries and new European Union members (the Czech Republic, Hungary, and the Slovak Republic), had quite modest fiscal responses to the pandemic.

Regressions confirm the sense from the figure that debt does not appear to have been destiny when it came to the pandemic fiscal response. Table 1 reports the results of simple cross-section regressions of the size of the early fiscal response on the 2019 debt-to-GDP ratio, with and without various control variables. Column 1 shows that in the most basic specification, the coefficient on the debt ratio actually enters positively; that is, countries with higher initial debt levels undertook more aggressive fiscal expansion. However, the standard error is so large that the two-standard-error confidence band encompasses both positive and negative values. Column 2 shows that the same pattern holds when net debt is used in place of gross debt.

Table 1. Relationship between Early Pandemic Fiscal Packages and Debt Ratios

Sample	*(1)* Full	*(2)* Full	*(3)* No Japan	*(4)* No Poor	*(5)* Full	*(6)* Full
Explanatory variable						
Gross debt/GDP	0.006		–0.001	0.005	0.007	0.020
	(0.009)		(0.013)	(0.009)	(0.008)	(0.007)
Net debt/GDP		0.001				
		(0.008)				
COVID-19 deaths					–0.002	
					(0.002)	
S&P rating						0.472
						(0.093)
Constant	4.745	5.276	5.108	5.541	4.981	–8.558
	(0.815)	(0.610)	(0.966)	(0.789)	(0.931)	(2.693)

Source: Author's calculations.

Notes: The dependent variable is the size of early pandemic fiscal packages (as a percentage of 2019 GDP), where a positive value corresponds to fiscal expansion. The gross and net debt ratios are as of the end of 2019 and are measured in percentage points; COVID-19 deaths are as of April 30, 2020; and the S&P rating is as of the end of 2019. The "No Poor" sample excludes Greece, Hungary, Mexico, Poland, the Slovak Republic, and Turkey. The standard errors reported are heteroscedasticity-consistent (Eicker-White) standard errors.

Column 3 shows that the lack of a relationship remains even if one takes out Japan, the most noticeable outlier, though the sign of the point estimate does flip to the expected negative relationship.[6] Column 4 shows that the lack of relationship also holds if one considers only relatively wealthy countries (and so excludes the six countries with the lowest GDP per capita in the sample—Greece, Hungary, Mexico, Poland, the Slovak Republic, and Turkey).

Column 5 shows that controlling for the initial severity of the COVID-19 outbreak in a country also does not reveal a negative effect of debt. I include cumulative COVID-19 deaths per 100,000 people as of April 30, 2020, on the assumption that countries with a worse outbreak would take more aggressive fiscal action for a given debt level.[7] Interestingly, COVID-19 deaths enter negatively (but insignificantly), and gross debt continues to enter positively (but again very insignificantly).

I also try including a direct measure of market access as a predictor of the early fiscal response to the pandemic. In particular, I include a country's

6. Excluding other outliers, such as Greece and the United States, also has little impact on the estimates.

7. The data on deaths come from Our World in Data, "Covid-19-data," https://github.com/owid/covid-19-data/blob/master/public/data/ecdc/total_deaths_per_million.csv, accessed February 11, 2021.

S&P sovereign bond rating as of the end of 2019 in the regression that also includes the gross debt-to-GDP ratio.[8] Column 6 of table 1 shows that the coefficient on the S&P rating is strongly positive and highly statistically significant. The coefficient estimate implies that a country with a rating three points higher (say AA versus A) is predicted to have an early pandemic fiscal package (as a percent of GDP) that is 1.4 percentage points higher.[9] The coefficient on the debt ratio remains positive and is now marginally significant.

That market access is a much more important determinant of the fiscal response to the pandemic than initial debt suggests an important change from the 2008 financial crisis. Countries in 2020 appear to have been constrained in their fiscal choices not by ideas related to debt and deficits but by their ability to borrow. This could suggest an evolution of economic ideas away from unwarranted concern about debt levels in times of stress. However, it could also reflect the unique terror engendered by the pandemic and countries' desire to combat it. Only time, and the next crisis, will tell.

II. Evaluating the US Fiscal Policy Response

The data on the size of fiscal responses to the pandemic across countries show that the US response was nothing short of enormous. But was it well conceived and appropriate for the unique conditions of a pandemic recession? Does it appear to have been effective?

II.A. Differences between Ordinary and Pandemic Recessions

The first step in evaluating the desirability of recent fiscal measures is thinking about how a pandemic recession differs from an ordinary recession. Most recessions involve a decline in aggregate demand, precipitated by a variety of factors, such as contractionary monetary policy, financial distress, or falls in consumer and business confidence. The pandemic recession also

8. The S&P data are from S&P Global, "Browse Ratings by Practice," https://www.standardandpoors.com/en_US/web/guest/entity-browse, accessed September 5, 2020. I convert it to a numerical scale, where AAA corresponds to 30, AA to 27, A to 24, and so on. Pluses and minuses move the value up or down one unit.

9. My findings about predictors of the fiscal response echo those Benmelech and Tzur-Ilan (2020) derived using a different sample of countries and a different approach to measuring fiscal responses. Balajee, Tomar, and Udupa (2020), Hosny (2021), and Apeti and others (2021) also examine the cross-country evidence concerning the fiscal response to the pandemic. One interesting finding from Apeti and others (2021) is that an alternative measure of fiscal space—a lower ratio of government debt to tax revenues—is associated with a larger fiscal policy response to the pandemic.

involved a large fall in demand. The high-frequency data on restaurant reservations, plane flights, and trips to retail establishments from last March show that consumers responded to emerging news about the virus by hunkering down—even before shelter-in-place orders required it.[10] Very high saving rates in the United States suggest that demand remained low throughout 2020. But that is where the similarity between ordinary and pandemic recessions stops.

DESIRABLE AMOUNT OF STIMULUS One difference involves the appropriate amount of aggregate demand stimulus. In an ordinary recession, a key role of policy is to try to get aggregate demand up any way possible. The goal is to fill the hole in demand and thus return output to full employment quickly. But a pandemic thrives on human interaction, and hence on economic activity. Even if fiscal policy could counteract the decline in aggregate demand caused by virus fears and uncertainty and thus maintain full employment, policymakers should not want to do so. Certain activities—indoor dining at restaurants, cruise travel, concerts, conventions, and sporting events—simply cannot occur safely during a pandemic. The goal of policy during a pandemic is to stimulate only as much production and employment as can happen relatively safely.

A related point is that what happens during a pandemic recession depends crucially on the public health situation. The course of the virus determines how much and what can be produced safely. It also plays a key role in demand. Without effective public health measures, aggregate demand stimulus would likely cause the virus to surge. This, in turn, would cause private demand to collapse, countering any benefits of the fiscal stimulus.

UNEQUAL IMPACT A second difference between ordinary and pandemic recessions involves the inequality in harm to different types of workers. While the effects of any recession tend to be unequal, the effects of a pandemic recession are uniquely so. Some workers, particularly nonmedical professionals in the service sector, are able to switch easily to working from home. Assuming demand for their remote services does not decline substantially, these workers are likely to experience relatively modest increases in unemployment during a pandemic. And, to the degree that people prefer the flexibility provided by working from home and not commuting, the benefits of working may even increase for such workers during a pandemic.

On the other hand, workers in sectors particularly affected by the pandemic, such as hospitality and brick-and-mortar retail, are likely to

10. See, for example, Goolsbee and Syverson (2021) and Chetty and others (2020).

be severely harmed. Whether because demand dries up or because of shutdown orders, workers in these sectors are likely to experience prolonged unemployment. Workers in essential sectors, such as health care or food manufacturing, who cannot work remotely are unlikely to become unemployed, but their jobs become riskier or less pleasant. As a result, they are also particularly harmed.

DOES STIMULUS FLOW THROUGHOUT THE ECONOMY? A third difference is that the benefits of aggregate demand stimulus do not flow throughout the economy during a pandemic recession. In an ordinary recession, it is not necessary to target aggregate demand stimulus to the particular sectors or people affected. For example, if residential construction declines, it is not necessary to focus on measures closely tied to residential construction. Any measure that stimulates demand will cause an increase in income that will flow through to increased demand for housing and so help the construction sector. This is much less true in a pandemic recession. Because some sectors cannot operate safely in a pandemic, general demand stimulus will do little to help unemployed workers in those sectors. The usual knock-on effects behind a traditional Keynesian multiplier—spending in one area flows to spending throughout the economy—fail to operate when part of the economy is shut down.[11]

IMPLICATIONS The unique characteristics of a pandemic recession imply that fiscal policy during a pandemic should be geared much more toward helping those who are directly harmed rather than toward increasing aggregate demand more generally. That is, it should be aimed at providing social insurance rather than broad stimulus. As we formalize and discuss in Romer and Romer (2021), the sensible role of policy during a pandemic is to provide people with the compensation they would have received if they had been able to insure themselves against the effects of a pandemic. Such targeted aid should ideally compensate not only those who become unemployed but also those who remain employed but at high risk of exposure because of the essential nature of their jobs. Directing aid to those affected deals directly with the problem of grossly unequal harms from the pandemic recession. It also deals with the problem that general stimulus does not flow throughout the economy during a pandemic.

Of course, to the extent that output during the pandemic is below the level that can be produced safely because of an aggregate demand shortfall, broad fiscal stimulus would be appropriate and desirable. Another benefit of

11. This point is formalized in the model of Guerrieri and others (2020). They also show that shutdowns can themselves have aggregate demand consequences.

Table 2. Deficit Impact of US Pandemic-Related Legislation

Provision	*Impact on deficit ($ billions)*
Enhanced unemployment benefits	748
Direct assistance to state and local governments[a]	597
Health care spending[b]	629
Direct payments to households	870
Paycheck Protection Program	808
Other loan and grant provisions	232
Other spending provisions[c]	890
Tax reductions	426
Total	5,200

Source: Author's calculations.

Notes: Data come from the Committee for a Responsible Federal Budget (CRFB) Covid Money Tracker (https://www.covidmoneytracker.org/explore-data/interactive-table), accessed April 24, 2021. The numbers were corroborated where possible using CBO documents.

[a]In addition to the $150 billion provided by the CARES Act for the Coronavirus Relief Fund and the $362 billion provided by the American Rescue Plan for the Coronavirus State and Local Fiscal Recovery Funds, I also include the $85 billion for Medicaid Matching Funds Increase provided by the Families First Act and extended by other acts.

[b]Of the $697 billion CRFB includes in health spending, $67.8 billion is more sensibly categorized as other spending because it is largely transfers to households. In particular, $22.8 billion is for government coverage of COBRA payments and $45 billion is for ACA subsidies.

[c]Other spending provisions include both the $401 billion CRFB reports in this category and pieces from other categories (such as the transfer components of health spending) that I subtract to get my preferred breakdown.

targeting aid only to those directly harmed is that such households are likely to spend the aid and thus contribute to overall demand. At the same time, because targeted fiscal support is more likely to be spent on necessities such as rent and food than general stimulus, it is less likely to encourage consumption and production that is unsafe, such as travel or entertainment.

II.B. Particular Fiscal Measures

With these general principles in mind, let me turn to a high-level evaluation of the fiscal measures taken in the United States in response to the pandemic. Table 2 shows the major components of the US fiscal response. Roughly 14 percent of the $5.2 trillion the United States allocated to pandemic aid went to expanded unemployment insurance (UI), and another 17 percent went to onetime direct payments to households (the so-called economic impact payments or stimulus checks). Another 16 percent went to the Paycheck Protection Program, which provided forgivable loans to small businesses if they maintained payrolls. About 11 percent went to aid to state and local governments. The remaining 42 percent of the total budgetary impact was attributable to an array of temporary spending increases and tax cuts. Of the miscellaneous spending, roughly $629 billion (or about 12 percent of

the total fiscal impact) went to public health measures, such as paying for COVID-19 care, vaccine development, and testing.

EXPANDED UNEMPLOYMENT INSURANCE The expansion of unemployment insurance was clearly appropriate to the unique circumstances of the pandemic. Pandemic-related legislation expanded the coverage of the unemployment insurance program to include workers such as Uber drivers and the self-employed. As discussed by Ruffini and Wozniak (2021), the ratio of those receiving benefits to the total number unemployed (the recipiency rate) was only about 30 percent in the twenty-five years before the pandemic. Thanks to the recent emergency measures, the recipiency rate has risen to close to 100 percent in the last year. In addition to covering workers who previously did not qualify for UI, emergency measures increased both the generosity of benefit payments and their duration. The increase in payments, particularly the across-the-board extra $600 per week provided by the Coronavirus Aid, Relief, and Economic Security (CARES) Act, raised replacement rates to well over 100 percent for unemployed low-wage workers (Ganong, Noel, and Vavra 2020).

The expanded unemployment insurance is firmly in the social insurance branch of fiscal policy. It provides aid to those directly affected by the pandemic. Increased replacement rates are also appropriate in a recession where many jobs are forbidden for public health reasons. However, a policy resulting in replacements rates in excess of 100 percent was clearly designed to accomplish additional goals, such as poverty reduction or aggregate demand stimulus.

As aggregate demand stimulus, the expanded unemployment insurance appears to have been quite effective. One heartening development during the pandemic has been the burgeoning of economics research on the effects of the pandemic itself and the policy response. Many of these studies use innovative, high-frequency data from private sector sources. For example, using proprietary bank account records, Farrell and others (2020) estimate that the spending of benefit recipients increased $0.73 for every $1 of additional benefits. Thus, to the degree that aggregate demand stimulus was appropriate during the pandemic, the UI expansion was a cost-effective way to provide it.

Examining the pandemic fiscal response through the lens of social insurance leads naturally to consideration of government-provided hazard pay. In addition to those who lose their jobs, people who remain employed as frontline essential workers are also directly economically affected by the pandemic. At the same wage as before, the benefits of working are reduced—perhaps substantially so—by the risks of social contact and the

unpleasantness of protective measures. Notions of fairness and a need to maintain a fully staffed essential sector suggest that extra payments to frontline essential workers are desirable. To the degree that such payments are called for because the government's provision of unemployment insurance reduces frontline workers' incentives to continue working, it is natural for the government (rather than private employers) to pay them.

There have been pilot programs and proposals for hazard pay during the pandemic (Kinder, Stateler, and Du 2020). The CARES Act allowed state and local relief funds to be used to cover limited programs for hazard pay.[12] A number of states, including Pennsylvania, Vermont, Louisiana, Maryland, and New Hampshire, set up such programs. The Health and Economic Recovery Omnibus Emergency Solutions (HEROES) Act, passed by the House of Representatives in May 2020, included a $200 billion fund for "pandemic premium pay." This provision was removed before a greatly changed and slimmed down relief measure was passed in December 2020. The failure to include a substantial hazard pay program is an important missed opportunity in the fiscal response to the pandemic.

STATE AND LOCAL FISCAL RELIEF Aid to state and local governments is another type of targeted fiscal support. State governments faced substantial increases in spending because of the pandemic.[13] The sources ranged from the obvious increase in public health expenditure and funds for the switch to online education, to the less obvious spending to stem the spread of the virus through the homeless population and the additional costs of providing state services remotely or in person with additional safety precautions. At the same time, state tax revenues dropped at least somewhat as unemployment rose (Auerbach and others 2020). Because most states have balanced budget requirements, some have already been forced to cut other types of spending and raise taxes, and many others are on the verge of doing so.

Sensible public policy suggests that useful state spending should not be cut simply because states cannot borrow to smooth over temporary emergency expenditures or falls in revenue. Transfers from the federal government, which can borrow, help states to maintain services and employment during a crisis. This spending can be thought of as another type of social insurance. It prevents cuts in state spending that citizens would likely

12. See "Department of the Treasury, Coronavirus Relief Fund for States, Tribal Governments, and Certain Eligible Local Governments," *Federal Register*, January 15, 2021.

13. See National Conference of State Legislatures, "State Fiscal Responses to Coronavirus (COVID-19)," https://www.ncsl.org/research/fiscal-policy/state-fiscal-responses-to-covid-19.aspx.

have wished to insure against had they contemplated the possibility of a pandemic.

Such transfers also provide some of the most cost-effective aggregate demand stimulus. Chodorow-Reich and others (2012) found that the state fiscal relief in the 2009 American Recovery and Reinvestment Act provided positive employment benefits both inside and outside the public sector at the remarkably low cost per job of just $27,000. The study also looked at the behavior of rainy-day funds following the 2009 state fiscal relief and found no evidence that the transfers were saved—contradicting a common fear expressed about such payments.

One of the earliest pandemic fiscal measures, the Families First Coronavirus Response Act, included an increase in the Medicaid matching percentage, which is a form of state fiscal relief. The CARES Act, passed later in March 2020, included about $150 billion of direct payments to state and local governments. The American Rescue Plan Act, passed in March 2021, included another $362 billion of such funds. In addition to maintaining state services and helping to deal with the effects of the pandemic, these funds are likely to have a substantially expansionary impact on aggregate demand in 2021. Whether this expansionary impact is desirable or not will depend on the evolution of the public health situation.

SPENDING ON PUBLIC HEALTH Spending on public health measures accounts for about $629 billion of the $5.2 trillion spent on pandemic relief. Given the widespread infections with COVID-19 and the large number of deaths, this spending was clearly necessary and valuable. Such spending was also consistent with the unique nature of the pandemic recession. Because both the recovery of demand and the safe pace of economic growth depend on getting the virus under control, it was imperative to take aggressive action on public health measures. Indeed, it is hard to imagine that more should not have been spent. Take, for example, what has clearly been one of the public health triumphs: vaccine development. The somewhat unfortunately named Operation Warp Speed used direct government spending and guaranteed sales contracts to encourage private pharmaceutical companies to develop vaccines. A number of highly effective vaccines were available within a year of the recognition of the virus—a record for vaccine development. However, the painfully slow progression from vaccine discovery to actual inoculation strongly suggests that more funding and effort was needed to set up effective and rapid distribution programs. Much larger capital expenditures for mobile refrigeration units and production facilities, along with free training for emergency providers, could likely have gotten the vaccines into many more arms much faster.

The United States has also failed to establish a truly effective virus monitoring program. Though testing has increased markedly, we still conduct fewer tests per one thousand people than many other rich countries. For example, Denmark currently conducts about nine times as many tests per person as the United States; the United Kingdom about seven times as many.[14] The United States is even worse at genetic sequencing of cases to identify variants and patterns of transmission. Iceland sequences about 60 percent of all cases; the United States sequences less than 1 percent. We currently rank thirty-second in the world for sequences completed per one thousand COVID-19 cases.[15] As a result, we know much less about emerging variants and where transmission is coming from. Based on the experience of other countries, we could have done much better at getting the virus under control if we had had a more effective and wider-reaching public health response.

The failure to spend adequately on public health measures may reflect, in part, policymakers' failure to realize the fundamental difference between a pandemic recession and an ordinary recession. Thinking of the current recession as just a replay of the Great Recession, with a virus in place of a global financial crisis, may have led policymakers to focus too much on dealing with the economic fallout of the pandemic and not enough on combating the root cause. Of course, a refusal to follow scientific evidence and advice also surely played a role.

ONETIME STIMULUS PAYMENTS Roughly $870 billion of the budgetary impact of the COVID-19 fiscal response came from onetime stimulus payments. The payments went to everyone below a certain income threshold. For example, the original checks of $1,200 per adult authorized in the CARES Act went to married couples earning up to $198,000. Because of their broad reach, the payments had at least some impact on reducing inequality temporarily, and they surely gave many households a much-needed boost at a difficult time.

What is not to like about the stimulus payments? The main drawback is that the help is very poorly targeted. Most of the money went to people who remained employed during the pandemic. The $1,200 checks (followed by $600 and $1,400 checks) were surely helpful to those hurt by the pandemic, but they were not nearly enough to truly hold life together for the most

14. Our World in Data, "Daily COVID-19 Tests per Thousand People," https://ourworldindata.org/grapher/full-list-daily-covid-19-tests-per-thousand.

15. The statistics come from COVID CG, "Global Lineage Surveillance," https://covidcg.org, accessed March 15, 2021, which use data from GISAID.

affected. A related problem goes back to the idea that general stimulus does not flow to those in need during a pandemic recession. Ordinarily, anything that raises aggregate demand in a recession eventually helps unemployed workers throughout the economy. But during a pandemic, general stimulus cannot help workers in sectors that remain closed or greatly restricted.

Another potential concern about the economic impact payments is that they could stimulate output and employment beyond the safe level. Particularly if the payments were spent on risky items like travel or indoor restaurant dining, the payments could exacerbate the pandemic. Interestingly, at least in the case of the initial round, the COVID-19 stimulus payments seem unlikely to have stimulated aggregate demand excessively. Coibion, Gorodnichenko, and Weber (2020) conducted a survey of a large sample of recipients. Most respondents said that they saved the payment or used it to pay down debt. Only about 15 percent of people surveyed report that they mostly spent their payment.[16] While this small effect is perhaps desirable from a disease-control viewpoint, it also suggests that the payments were not particularly valuable as a relief or recovery mechanism.

PAYCHECK PROTECTION PROGRAM A novel and very substantial component of the fiscal policy response in the United States was the Paycheck Protection Program, or PPP. The program shares some features with the widely admired *Kurzarbeit* program in Germany. By providing loans that morphed into grants to small businesses that used the money primarily to maintain payrolls, the program was designed to preserve worker-firm matches. Workers received most, if not all, of their pay without becoming officially unemployed or applying for unemployment insurance.

One peculiarity of the program is that the forgivable loans were only available to firms with fewer than three hundred employees. Though one can tell stories that might lead policy to focus on firms facing borrowing constraints, being above or below three hundred employees is surely an extremely crude proxy on which to build an $808 billion program. Related to this point, there is some evidence that PPP loans tended to go toward firms that already had banking relationships (and so perhaps already had access to credit), rather than to the neediest ones (Liu and Volker 2020).

At least two studies have used the fact that eligibility was discontinuous to try to measure what the program accomplished (Chetty and others 2020; Autor and others 2020). Both find that employment declined less at firms

16. Using high-frequency data, Chetty and others (2020) found evidence of an immediate impact of the payments on consumer spending. However, the estimates do not show whether the effects were quickly undone or more persistent.

just below the eligibility cutoff than at those just over. However, the difference was quite small. Indeed, under reasonable assumptions, the implied cost per job (and hence per employee-firm match preserved) was very high—$224,000 in Autor and others (2020) and $377,000 in Chetty and others (2020). And, as pointed out by Hubbard and Strain (2020), it is too early to know if the matches were truly preserved or if workers will eventually take other jobs or be laid off once the program ends.

One reason why the cost per job preserved has been so high is that firms could receive grants to maintain matches that would have been maintained without government support. Another source of the high program cost is that the replacement rates and the maximum covered salary expenses were substantially higher than under conventional unemployment insurance. As discussed in Romer and Romer (2021), in a social insurance framework, high-income earners would likely choose to self-insure rather than purchase pandemic insurance. Thus, covering their wages during unemployment is not a sensible component of a public insurance program. The bottom line is that the PPP was an interesting and noble experiment, but it was problematic on many levels.

Overall, the fiscal response to the pandemic in the United States runs the gamut from highly useful and appropriate to largely ineffective and wasteful. Spending on programs such as unemployment compensation and public health was exactly what was called for by the unique nature of the pandemic recession. Spending on broad-based payments and other general stimulus measures was much less useful in a recession where the impacts were highly unequal and the Keynesian multiplier was likely substantially reduced by lockdowns.

III. Will the Pandemic Fiscal Response Have Repercussions for the Future?

The preceding analysis examined the more immediate appropriateness of various pandemic fiscal measures. I want to turn now to possible longer-term repercussions. Will the extraordinary fiscal measures taken during the pandemic have consequences for the future?

III.A. Is the Economy Likely to Boom?

One area of current discussion is whether the economy is likely to recover rapidly following the passage of the Biden administration's American Rescue Plan Act in March 2021. Considering the unique nature of the pandemic recession, the answer depends most heavily on what happens

to the virus. If the vaccination program is successful and COVID-19 cases moderate further, the economy is likely to do well. If vaccinations flag or become less effective as new variants emerge, recovery is likely to slow or stall.

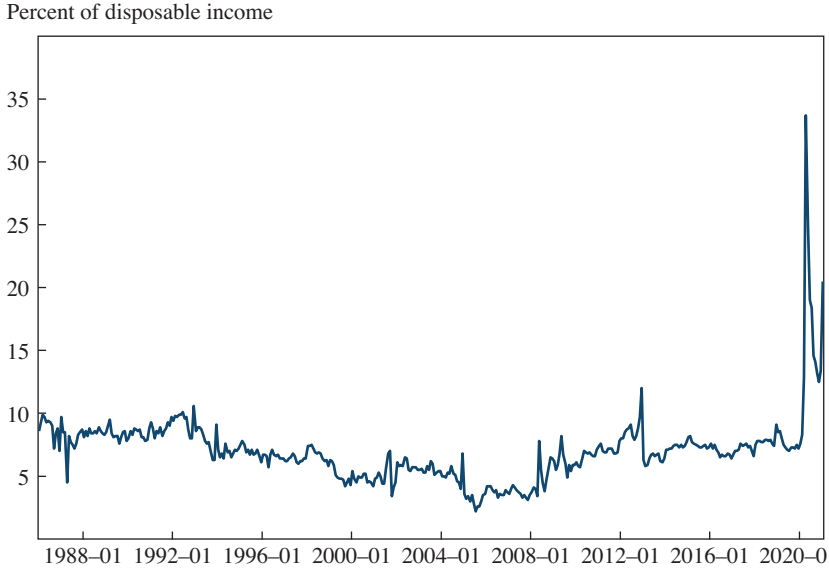
The American Rescue Plan Act provides $155 billion in health care spending, of which about $60 billion is focused more narrowly on vaccine and treatment development, vaccine distribution, and COVID-19 testing, tracing, and monitoring.[17] As discussed in section II, this funding is surely valuable and likely to speed control of the virus. By doing so, this aspect of the bill should help accelerate recovery. The unemployment insurance enhancements, stimulus payments, and state fiscal relief also included in the act are likely to provide a substantial fillip to aggregate demand. This demand increase, combined with greater virus control, could generate substantial increases in output and employment over the next two years.

Probably more important than the direct stimulus from the most recent pandemic fiscal package is the accumulated savings of American households. As can be seen in figure 3, the personal saving rate in the United States has been two to three times higher than normal since the virus emerged. Between this increased saving and the rise in stock prices, the level of financial assets of households and nonprofit organizations has risen almost $10 trillion since the fourth quarter of 2019. In recent work, Brunet (2019, 2021) suggests that the best parallel to the current situation is the buildup of savings during World War II. The combination of rationing, increased earnings, and production limitations forced households to save tremendously during the war. Following the war, households increased personal consumption just as tremendously. Using county-level data, Brunet (2019) finds that residential investment and related spending on household durables following the war increased significantly more in counties where the buildup of savings was larger. It is very possible that we will see a similar burgeoning of consumer demand driven by accumulated savings following the pandemic, once it is safe for people to shop and produce the goods that consumers desire.

There are, of course, factors that could hold back this surge of spending. It is possible that living through a pandemic will lead households to

17. The estimates are from the Committee for a Responsible Federal Budget (CRFB) COVID Money Tracker. I exclude the $67.8 billion for COBRA coverage and ACA subsidies from the CRFB's estimate of health spending in the American Rescue Plan Act. The act provides another $10 billion to purchase, manufacture, and distribute critically needed medical supplies and equipment under the authority of the Defense Production Act.

Figure 3. Personal Saving Rate in the United States since 1986



Source: Bureau of Economic Analysis, National Income and Product Accounts, table 2.6, accessed March 18, 2021.

Note: The data are for the period January 1986–January 2021.

permanently increase their precautionary saving. Likewise, households that were allowed to miss rent or loan payments will face higher debt loads as a result of the deferrals and so may be less able to spend. Many workers are also facing difficult transitions. For example, workers who had to leave the labor force to take care of children will likely face difficulties returning to the labor market at their previous wage. And the pandemic may lead to permanent changes in the sectoral composition of the economy that will require workers to retrain and find new employment. All of these are factors that could mute some of the surge in consumption that would otherwise occur. Nevertheless, I strongly suspect that the positive forces will outweigh these negative ones.

While output and employment are likely to increase markedly as the virus recedes and consumer demand increases, it is important to note just how far employment in the United States is below its normal trend level. Payroll employment in February 2021 was 9.5 million below its prepandemic high in February 2020. This is greater than the peak loss of jobs in the 2008 recession. Taking into account the normal trend growth of employment

adds at least another million jobs that are currently missing from the US economy. As a result, the United States needs to grow rapidly for a sustained period to heal the labor market, and it can do so without pushing up against the capacity constraints of the economy.

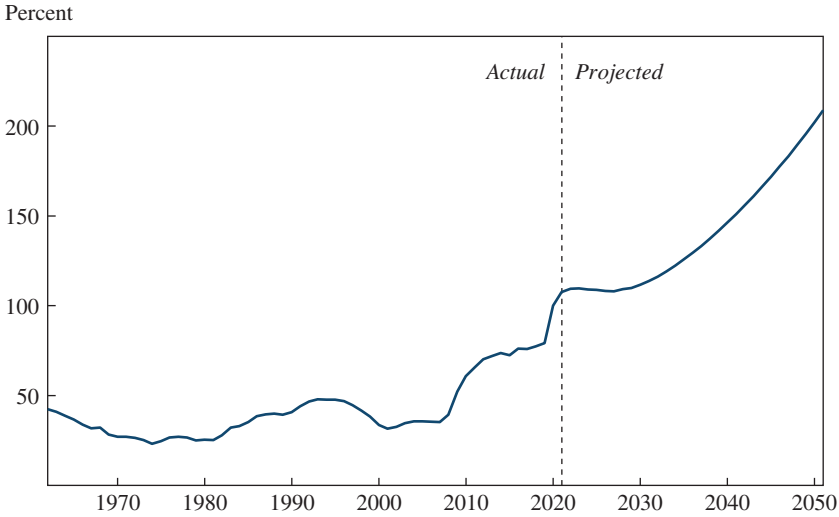
III.B. Reduction in Fiscal Space

A longer-term repercussion of the enormous fiscal response to the pandemic in the United States involves the increase in government debt. Figure 4 shows actual and projected federal debt-to-GDP ratios from 1962 to 2051. The data are from the Congressional Budget Office, updated to include the CBO's estimates of the deficit impact of the American Rescue Plan Act (CBO 2021).[18] The debt-to-GDP ratio stood at 79 percent at the end of fiscal year 2019 (before the pandemic had emerged). It is projected to reach 110 percent by the end of fiscal 2023 (when the spending from the American Rescue Plan has had its full impact). This sharp rise is obviously related not just to the deliberate fiscal response to the pandemic but also to the operation of automatic stabilizers. At the same time, sharp falls in current and expected interest rates brought about by the pandemic have been a factor pushing the debt ratio in the opposite direction (because lower interest rates reduce the cost of debt service). Nevertheless, the net result has been a substantial rise in the debt-to-GDP ratio, and thus a reduction in our fiscal space.

The high debt loads are unlikely to precipitate any kind of fiscal crisis in the United States. Demand for US government debt remains as strong as ever. But that does not mean that the rise in the debt ratio is costless. I fear it could lead to inaction on a number of national priorities. As discussed in section I, previous research shows that the fiscal response to financial distress in the postwar period has depended strongly on the initial debt-to-GDP ratio: countries throughout the OECD behaved as if they were constrained by high debt when responding to a financial crisis. While this does not seem to have been true of the response to the pandemic, the historical behavior could reemerge as the pandemic wanes. As the United States recovers and policymakers seek to tackle issues such as climate change, crumbling infrastructure, and persistent poverty, they may find increased opposition to further spending. Thus, one potential legacy of the extraordinary fiscal actions to fight the pandemic may be that the country fails to deal with other pressing needs.

18. I am grateful to Alan Auerbach for providing these data.

Figure 4. Actual and Projected US Federal Debt-to-GDP Ratio



Sources: CBO; calculations by Alan Auerbach.

Notes: The data through 2020 are from the CBO, "Budget and Economic Data: Historical Budget Data," https://www.cbo.gov/data/budget-economic-data, accessed February 2021. The data after 2020 are from calculations by Alan Auerbach, which take the long-term projections from CBO (Long-Term Budget Projections, March 2021) and adjust them for CBO's estimates of the ten-year costs of the American Rescue Plan Act (CBO 2021). The estimates assume no macroeconomic feedback from the act, and so hold the paths of GDP and the debt service/debt ratio constant.

This possibility puts the enormous size and significant flaws of the US fiscal response in a somewhat harsher light. Though much that was done was useful and unquestionably necessary, some was misguided and wasteful. If something like the nearly $1 trillion spent on stimulus payments that did little to help those most affected by the pandemic ends up precluding spending $1 trillion on infrastructure or climate change in the next few years, the United States will have made a very poor bargain indeed.

ACKNOWLEDGMENTS I thank Alan Blinder, Janice Eberly, David Romer, and James Stock for helpful comments, and Alan Auerbach and Gillian Brunet for data.

References

Apeti, Ablam Estel, Jean-Louis Combes, Xavier Debrun, and Alexandru Minea. 2021. "Did Fiscal Space Influence Covid-19's Fiscal Response?" *Covid Economics: Vetted and Real-Time Papers* 74 (March 30): 71–93.

Auerbach, Alan J., William G. Gale, Byron Lutz, and Louise Sheiner. 2020. "Effects of COVID-19 on Federal, State, and Local Government Budgets." *Brookings Papers on Economic Activity*, Fall, 229–70.

Autor, David, David Cho, Leland D. Crane, Mita Goldar, Byron Lutz, Joshua Montes, William B. Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz. 2020. "An Evaluation of the Paycheck Protection Program Using Administrative Payroll Microdata." Working Paper. https://economics.mit.edu/files/20094.

Balajee, Anuragh, Shekhar Tomar, and Gautham Udupa. 2020 "COVID-19, Fiscal Stimulus, and Credit Ratings." *Covid Economics: Vetted and Real-Time Papers* 11 (April 29): 132–64.

Benmelech, Efraim, and Nitzan Tzur-Ilan. 2020. "The Determinants of Fiscal and Monetary Policies during the Covid-19 Crisis." Working Paper 27461. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27461.

Brunet, Gillian. 2019. "After the War: Wartime Saving and Postwar Housing Investment, 1946–1950." Working Paper. http://conference.nber.org/conf_papers/f127354.pdf.

Brunet, Gillian. 2021. "War, Pandemic, and Household Behavior: The COVID-19 Pandemic through the Lens of WWII." Work in progress.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Michael Stepner, and the Opportunity Insights Team. 2020. "The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data." Working Paper 27431. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27431.

Chodorow-Reich, Gabriel, Laura Feiveson, Zachary Liscow, and William Gui Woolston. 2012. "Does State Fiscal Relief during Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act." *American Economic Journal: Economic Policy* 4, no. 3: 118–45.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2020. "How Did U.S. Consumers Use Their Stimulus Payments?" Working Paper 27693. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27693.

Congressional Budget Office (CBO). 2020. "The Effects of Pandemic-Related Legislation on Output." Washington. https://www.cbo.gov/publication/56597.

Congressional Budget Office (CBO). 2021. "Estimated Budgetary Effects of H.R. 1319, American Rescue Plan Act of 2021." Washington. https://www.cbo.gov/publication/57056.

Farrell, Diana, Peter Ganong, Fiona Greig, Max Liebeskind, Pascal Noel, and Joseph Vavra. 2020. "Consumption Effects of Unemployment Insurance during the Covid-19 Pandemic." Working Paper. Washington: JPMorgan Chase Institute.

https://www.jpmorganchase.com/institute/research/labor-markets/unemployment-insurance-covid19-pandemic.

Ganong, Peter, Pascal Noel, and Joseph Vavra. 2020. "US Unemployment Insurance Replacement Rates during the Pandemic." *Journal of Public Economics* 191, article no. 104273.

Goolsbee, Austan, and Chad Syverson. 2021. "Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020." *Journal of Public Economics* 193, article no. 104311.

Gravelle, Jane G., and Donald J. Marples. 2021. *Fiscal Policy and Recovery from the COVID-19 Recession*. Washington: Congressional Research Service. https://crsreports.congress.gov/product/pdf/R/R46460.

Guerrieri, Veronica, Guido Lorenzoni, Ludwig Straub, and Iván Werning. 2020. "Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages?" Working Paper 26918. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26918.

Hosny, Amr. 2021. "The Sooner (and the Smarter), the Better: COVID-19 Containment Measures and Fiscal Responses." *Covid Economics: Vetted and Real-Time Papers* 73 (March 23): 137–57.

Hubbard, Glenn, and Michael R. Strain. 2020. "Has the Paycheck Protection Program Succeeded?" *Brookings Papers on Economic Activity*, Fall, 335–78.

Kinder, Molly, Laura Stateler, and Julia Du. 2020. *The COVID-19 Hazard Continues, but the Hazard Pay Does Not: Why America's Essential Workers Need a Raise*. Washington: Brookings Institution. https://www.brookings.edu/research/the-covid-19-hazard-continues-but-the-hazard-pay-does-not-why-americas-frontline-workers-need-a-raise/.

Liu, Haoyang, and Desi Volker. 2020. "Where Have the Paycheck Protection Loans Gone So Far?" Blog post, May 6, Liberty Street Economics, Federal Reserve Bank of New York. https://libertystreeteconomics.newyorkfed.org/2020/05/where-have-the-paycheck-protection-loans-gone-so-far.html.

OECD. 2009. "The Effectiveness and Scope of Fiscal Stimulus." In *Economic Outlook, Interim Report March 2009*. https://www.oecd-ilibrary.org/economics/oecd-economic-outlook-interim-report-march-2009/the-effectiveness-and-scope-of-fiscal-stimulus_eco_outlook-v2008-sup2-16-en.

Romer, Christina D., and David H. Romer. 2018. "Why Some Times Are Different: Macroeconomic Policy and the Aftermath of Financial Crises." *Economica* 85, no. 337: 1–40.

Romer, Christina D., and David H. Romer. 2019. "Fiscal Space and the Aftermath of Financial Crises: How It Matters and Why." *Brookings Papers on Economic Activity*, Spring, 239–313.

Romer, Christina D., and David H. Romer. 2021. "A Social Insurance Perspective on Pandemic Fiscal Policy: Implications for Unemployment Insurance and Hazard Pay." Working Paper. University of California, Berkeley.

Ruffini, Krista, and Abigail Wozniak. 2021. "Supporting Workers and Families in the Pandemic Recession: Results in 2020 and Suggestions for 2021." In the present volume of *Brookings Papers on Economic Activity*.

KRISTA RUFFINI
Georgetown University

ABIGAIL WOZNIAK
Federal Reserve Bank of Minneapolis

Supporting Workers and Families in the Pandemic Recession: Results in 2020 and Suggestions for 2021

ABSTRACT We review several spending programs designed to support Americans through the COVID-19 pandemic in 2020. We group these into programs designed to stabilize the labor market and facilitate its recovery and those that provided financial relief to households independent of their employment history. We review the extent to which these programs reached intended beneficiaries along with early evidence of program impacts. Overall, we find the programs were highly successful at delivering intended aid in 2020. Nevertheless, we identify common areas where programs could improve as support continues through 2021, and we discuss related needs that have so far received less attention from policymakers.

In 2020, the US economy experienced the sharpest contraction on record as shutdowns and behavioral changes to contain the SARS-CoV-2 virus rapidly took hold across the globe. Between April 2019 and April 2020, more than 20 million people—about 12 percent of the US labor force—lost employment. In the second half of 2020, the economy started to recover, but twelve months after the first US shutdowns, economic hardship remains

Conflict of Interest Disclosure: Krista Ruffini is a visiting scholar at the Federal Reserve Bank of Minneapolis. Abigail Wozniak is a labor economist at the Federal Reserve Bank of Minneapolis. She also holds unpaid advisory board positions at the McKinsey Institute for Black Economic Mobility and the Hennepin County Small Business Recovery task force. Other than the aforementioned, the authors did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this paper. They are currently not officers, directors, or board members of any organization with an interest in this paper. The views expressed here are the authors' and do not necessarily reflect those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

an acute concern. There are 9.7 million people unemployed as of March 2021, nearly twice the prepandemic level, and labor force participation has fallen by more than 3.8 million people. Though it is challenging to make comparisons to prepandemic data, one-third of households reported difficulty paying for usual expenses as of early January.[1] These rates of economic hardship are considerably higher among Black and Hispanic households and those with less than a college education.

The ongoing pandemic drives these economic challenges, even absent formal policy changes or significant local risk of infection (Goolsbee and Syverson 2021; Couture and others 2020; Chen, Qian, and Wen 2021). It is likely that the pandemic will continue to disrupt economic activity in the short- and medium-term, as vaccination is expected to continue through mid-2021 and the risk of infection from more contagious and possibly more severe COVID-19 variants increases.

Early in the pandemic, the US government launched a large, multifaceted policy response aimed at stabilizing US employment and protecting worker and household well-being from unexpected income losses. The main elements of this response were designed and enacted within seventy days of the first confirmed COVID-19 case in the United States and within fourteen days of most US shutdowns. One year later, it is an opportune time to examine this initial response and to consider the lessons it offers for the second year of combating pandemic-related economic disruption.

In this paper, we focus on two sets of policy responses enacted between March and December 2020. First, we consider programs intended to stabilize employment relationships and employment-based income: unemployment insurance (UI) supplements and the Paycheck Protection Program (PPP). Second, we review programs that aimed to support households' balance sheets, largely apart from their labor force participation. These include Economic Impact Payments (EIPs), Supplemental Nutrition Assistance Program (SNAP) expansions, and eviction moratoria. All together, these programs affected broad portions of the US workforce, on both the worker and firm sides, and substantially expanded the US safety net. We review the goals of these programs and provide a high-level assessment of whether they were met. We then discuss how lessons from this initial response should inform policy parameters going forward. While these programs have different goals and target populations, they were broadly intended to support Americans' financial well-being through the pandemic and were all enacted as onetime interventions or with specific expiration dates or caps. As such, continued

1. US Census Bureau, "Week 22 Household Pulse Survey: January 6–January 18," https://www.census.gov/data/tables/2021/demo/hhp/hhp22.html.

policy action is necessary if these programs are to continue in the medium term. We conclude with a discussion of several needs that were overlooked in the initial policy response.

Before turning to the specifics of each program, it is worthwhile to review the US fiscal response in 2020. First, the overall response was large. In the first six months of the crisis, Congress appropriated nearly $2.6 trillion in new agency spending and provided an additional $900 billion in tax relief, greater than the amount passed on fiscal support legislation over five years during the Great Recession (Council of Economic Advisers 2014).[2] As a share of GDP, the size of the US fiscal response ranks near the median of other OECD countries (Elgin, Basbug, and Yalaman 2020).

Second, assistance took many forms. Some components of relief expanded existing programs, such as more generous SNAP and UI payments. Others created entirely new programs, often administered at the state level, like Pandemic Unemployment Assistance (PUA). Other components were hybrid efforts that built upon existing structures and systems, like stimulus payments issued primarily through the tax system.

The initial response assumed a short and severe downturn. Some policies seem to be designed considering one of those assumptions more than the other. A belief that the contraction would be short motivated the PPP's efforts to preserve existing employment relationships, while the likelihood of a severe contraction motivated large UI supplements and eviction moratoria. Many components of this relief were intended to be temporary while public health officials developed an effective virus response. However, the pandemic continued and accelerated during summer and fall 2020, while many of the provisions in the early legislation expired. Moreover, some provisions were not tailored to the unique nature of the COVID-19 downturn but rather to an understanding of how programs behaved historically. For example, the fact that large numbers of workers are not covered by UI motivated covering these workers through new PUA payments.

I. The Reach of Stabilization and Financial Relief Efforts

In this section, we examine federal policy responses in two areas: those aimed at preserving prepandemic labor market activity and those aimed at preventing financial vulnerability. First, we examine programs that aimed

2. Throughout, we focus on the response in 2020 as provisions considered in early 2021 are too recent to evaluate. Our numbers do not include provisions for additional funds to these programs included in the $1.9 trillion American Rescue Plan Act in March 2021. We discuss relevant March 2021 policy updates in section II.

to stabilize income streams for workers and firms, thereby facilitating faster economic recovery: UI expansions and the PPP. Second, we consider programs that provided financial relief to households: EIPs, SNAP emergency allotments, and eviction moratoria. For each program, we briefly review its rationale, target population, and administrative design. We then summarize the early evidence on whether each program reached its target populations and met its goals. Table 1 overviews these programs. Although direct comparisons across such a diverse set of programs are not always possible, we present per recipient expenditures and total expenditures between March and December 2020 to give a sense of the scale of each program.

I.A. Policies to Stabilize the Labor Market and Support the Recovery

UNEMPLOYMENT INSURANCE Unemployment insurance (UI) aims to help individuals maintain consumption if they lose their job or are placed on temporary layoff due to changing business conditions or employer needs. It may also keep workers connected to the labor force through downturns and separations. Early in the pandemic, many job losses were classified as short-term layoffs. Over the following months, many of these layoffs became permanent job losses (Hedin, Schnorr, and von Wachter 2020) and overall, more than 9 million fewer people were working in December 2020 than in December 2019. Traditional UI benefits would have provided some support to these workers, but those who received benefits would have received only a fraction of their usual income and many workers would have been left out due to gaps in coverage.

Several changes to the UI system in 2020 expanded both eligibility and the generosity of payments in anticipation of a short and sharp downturn. First, in order to offset the income loss accompanying unemployment and support consumer spending, the Federal Pandemic Unemployment Compensation (FPUC) benefit provided an additional $600 a week to UI recipients between March 29 and July 25, 2020, as part of the Coronavirus Aid, Relief, and Economic Security (CARES) Act (Davies and Morton 2020). Before FPUC, the median weekly state benefit was $300, so FPUC tripled benefits for the typical worker, with about three-quarters of FPUC recipients receiving more in UI income than their previous earnings (Cortes and Forsythe 2020; Ganong, Noel, and Vavra 2020; Moffitt and Ziliak 2020).

The rationale for generous unemployment benefits was that providing a large wage subsidy early in the downturn would boost aid to those most in need, as well as help prevent a deeper or longer recession by supporting

Table 1. Summary of Policies to Stabilize the Labor Market and Provide Financial Support

Policy/Program	*Dates authorized (in 2020)*	*Actual recipients (millions)*	*Disbursement frequency*	*Average payment ($)*	*Total amount ($ billions, March–December 2020)*
Economic Impact Payments (cash payments)	April and December	174.7 households	Lump sum, twice	2,610	456.0
Supplemental Nutrition Assistance Program (food assistance)	March	22.6 households	Monthly	348	51.6
Unemployment insurance (benefit and duration extensions; cash payments)	FPUC: March; LWA: August; PEUC: March[a]	10.3 continuing claims	Biweekly or weekly	306 per week	122.5
Pandemic Unemployment Assistance (cash payments)	March	5.7 continuing claims	Biweekly or weekly		
Paycheck Protection Program (small business loan)	April and December	5.55 loans	Lump sum, three times	100,500	558.0
Eviction moratoria (deferral policy)	September	1.6 renters	In effect through March 2021	7,016	11.2

Source: Authors' compilations using data and sources as of January 2021.

Notes: EIPs include spring and winter 2020 estimated payments from the Tax Policy Center; SNAP amounts through September 2020 are from caseload data (USDA); UI amounts through September 2020 derived from quarterly claims and financial report (DOL); the PPP come from the SBA report (approvals through January 24, 2021); eviction moratoria estimated based on simulations from Reed and Divringi (2020), assuming 50 percent UI recipiency rate. FPUC is Federal Pandemic Unemployment Compensation, LWA is Lost Wages Assistance, and PEUC is Pandemic Emergency Unemployment Compensation.

a. Authorizing legislation specified end dates for these programs. FPUC payments authorized in March ended in July 2020. LWA authorized in August ended on or before October 2020. PEUC authorized in March added 24 weeks of additional payments after a worker exhausted regular payments through September 2021.

household spending. A side benefit of this approach was that minimizing person-to-person contact slows the spread of communicable diseases. Therefore, during the pandemic recession, encouraging those out of work to search for new employment could be counterproductive if labor market reentry accelerated the spread of the virus.

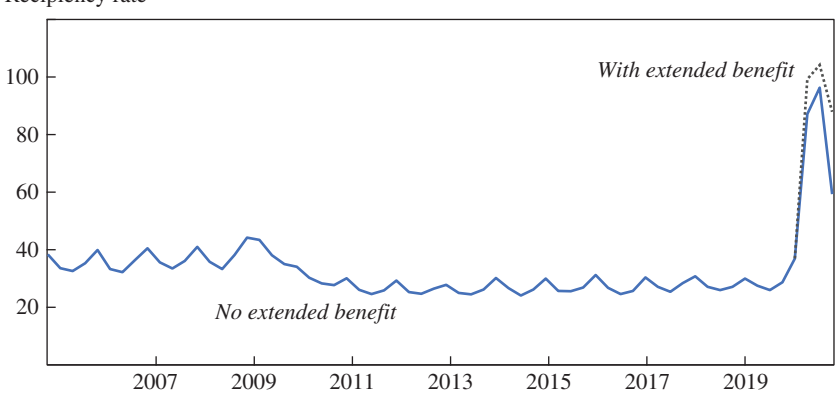
In August 2020, FPUC expired and was replaced with the Lost Wages Assistance (LWA) program (Davies and Morton 2020), which reduced the additional UI benefit to $300 and provided for an additional three to six weeks of payments. In contrast to FPUC, LWA was only available to recipients who were receiving at least $100 a week in other UI benefits, and workers with the lowest earnings were not eligible. In addition, LWA was a joint federal-state program: states had to apply for federal funding for a specified benefit duration and total program expenditures were limited by the Disaster Relief Fund. By the end of October 2020, all state LWA payments had expired.

Similar to previous downturns, Congress also extended the number of weeks that laid-off individuals could receive benefits, though the ultimate duration of weekly benefits varied by state and workers' filing history. Starting in March 2020, Pandemic Emergency Unemployment Compensation (PEUC) provided an additional thirteen weeks of benefits to these workers that was extended to twenty-four weeks under the Continued Assistance to Unemployed Workers Act of 2020. Broadly, these extensions meant that few, if any, workers who qualified for regular UI would have seen their benefits lapse in 2020, though the specifics vary by state and worker.

The CARES Act also created a new program, Pandemic Unemployment Assistance (PUA) to expand UI eligibility to workers who were ineligible for regular UI, including those with short work histories, those working as independent contractors ("gig workers"), and those who were self-employed. This population is likely a sizable share of unemployed workers: prior to 2020, only about 30 percent of unemployed workers were receiving UI benefits.[3] Like regular UI, states were charged with administering PUA, so each state had to develop a new program and establish replacement rates, maximum benefits, and verification rules. PUA was a large expansion of state UI systems, accounting for more than 35 percent of continuing claims by January 2021, as shown in table 1. Beyond the need to develop PUA systems, the large uptick in claims during spring 2020 overwhelmed systems and delayed payments. As a result, there was wide variation in when the

3. US Department of Labor, "Unemployment Insurance Data," https://oui.doleta.gov/unemploy/data_summary/DataSum.asp, accessed February 5, 2021.

Figure 1. Unemployment Insurance Recipiency Rate, 2005:Q1–2020:Q3



Source: US Department of Labor, Employment and Training Administration.

Notes: Solid line has only regular UI payments in the numerator; dashed line includes both regular and extended benefit payments. Data available at United States Department of Labor, "Unemployment Insurance Data," https://oui.doleta.gov/unemploy/data_summary/DataSum.asp.

first PUA payments were disbursed, ranging from March to June, from state to state.

The suite of changes in UI greatly expanded access to these benefits. One way to see this is through the increased ratio of the number of unemployment claims processed to the total number of unemployed people. Department of Labor (DOL) data show that this ratio, the recipiency rate, jumped from about 30 percent between 2005 and 2019 to 96 percent in the third quarter of 2020 (figure 1). The recipiency rate is an imperfect measure of benefits receipt among the unemployed for a number of reasons, and some data suggest receipt rates far below the published DOL rate (though still above historic levels).[4] Nonetheless, figure 1 indicates a significant expansion of UI access relative to historic patterns. Moreover, early research suggests FPUC increased spending among unemployed workers as intended, and spending fell when benefits were reduced under LWA

4. In the pandemic environment, three factors might inflate recipients (claims) relative to the measured unemployed, raising the recipiency rate: with waived job search requirements, not all UI recipients may be actively looking for work while claiming benefits; fraudulent claims raise claims relative to the unemployed; and misclassification of workers on temporary layoff will reduce measured unemployed. Survey data suggest imperfect delivery: in the earliest January Household Pulse Survey only about three-quarters of those who have applied for UI are currently receiving benefits.

(Farrell and others 2020, "Unemployment Benefit Boost"; Farrell and others 2020, *Consumption Effects*). In addition, given the lump-sum benefit and high replacement rate, FPUC increased income at the bottom of the distribution and reduced inequality (Cortes and Forsythe 2020). Although payment delays during the spring led households to sharply cut spending while waiting for benefits, consumption increased once they had received payments (Farrell and others 2020, *Consumption Effects*).

Early work examining the determinants of the 2020 labor market contraction finds that concern about virus spread and resulting low demand for in-person services drove high rates of joblessness, and thus changes to UI in 2020 were unlikely to have further reduced employment (Chetty and others 2020; Goolsbee and Syverson 2021). Consistent with this, analysis from the first several months of the pandemic finds that generosity of UI benefits—including the additional amounts paid under FPUC and LWA—did not significantly slow the recovery in 2020 (Altonji and others 2020; Bartik, Bertrand, and others 2020). Although it is still too early to fully examine the effect of longer benefit duration, evidence from the Great Recession suggests that extensions alone are unlikely to be a driver of the tepid labor market recovery (Boone and others 2021). It is important to note that these effects could change as the pandemic recedes, and the impacts of UI expansions on employment in 2021 could differ substantively from their impacts in 2020.

PAYCHECK PROTECTION PROGRAM The Paycheck Protection Program (PPP) distributed forgivable loans for compensation, business rent, mortgage, and utilities to small businesses that retained workers during the pandemic downturn. The goal of the PPP was to preserve labor market relationships. Although advocates took different views of whether the deeper rationale was to preserve jobs or firms, underlying this approach was the assumption that the pandemic contraction was temporary and unrelated to economic fundamentals.

The PPP aimed to support small businesses, generously defined as those with fewer than 500 employees in the initial authorization (later lowered to under 300 employees) and covered a broad range of entities, including nonprofits. Such businesses account for 47 percent of prepandemic employment (Hubbard and Strain 2020).[5] Federal funding for the program totaled $943 billion across three waves: two in April 2020 and a third in December (Liu and Volker 2020a; Strain 2020). The Small Business Administration

5. To further support the hard-hit food and accommodation sector, the size cap was applied on a per establishment basis for those firms.

(SBA) was charged with administering loans through its network of lenders, who were allowed to charge fees and had few obligations under the program. To facilitate timely disbursement of funds, the Federal Reserve developed a liquidity facility under its Section 13(3) authority that extended credit to eligible SBA lenders and accepted PPP loans as collateral (Liu and Volker 2020b). The PPP was modeled after Great Recession labor market interventions in Europe and utilized existing policy levers like the SBA loan network and the Federal Reserve's Section 13(3) powers, but it was ultimately a novel policy approach in the United States (Giupponi and Landais 2020).[6]

The PPP was a large program. As Autor and others (2020) note, April 2020 PPP funding was equivalent to two and a half months of total payroll for the roughly 60 million employees of US small businesses prepandemic. Despite media reports of problematic roll out, subsequent research has shown that PPP loans reached very large shares of eligible businesses in its first thirty days of operation (Autor and others 2020). Responses to the Census Bureau's Small Business Pulse Survey indicate that upward of 80 percent of businesses with five or more employees applied for PPP loans, and nearly all those that applied received a loan. Businesses with one to four employees had lower rates of PPP access, closer to 60 percent. These smallest businesses may have lacked banking relationships that facilitated PPP access or employee-owners of such businesses may have received support through PUA. In addition, such businesses have high exit rates even in non-pandemic times.[7] Given these numbers, the PPP was largely successful in terms of administration and reach to target firms. Though disbursement to the smallest firms and those serving communities of color was initially lower, access improved in the second round of the program (Fairlie and Fossen 2021).[8] This success is notable given that the PPP is essentially unique to the pandemic and is larger than the entire American Recovery and Reinvestment Act.

The program's ultimate goal was—depending on one's perspective—either to preserve jobs or to preserve businesses. Evidence from 2020 suggests generally modest or insignificant impacts on employment but

6. The US UI system allows for a more traditional short-time compensation program (STC) than the PPP, but it has very low utilization rates. At the peak of the pandemic, STC accounted for only about 1 percent of total UI claims (Krolikowski and Weixel 2020).

7. We are grateful to David Cho (Federal Reserve Board of Governors) for sharing these tabulations with us from his discussion at the December 2020 Center for Human Capital Studies conference at the Federal Reserve Bank of Atlanta.

8. Fairlie and Fossen (2021) also show that the smaller Economic Injury Disaster Loan (EIDL) program was distributed more equitably to the smallest businesses immediately.

potentially meaningful improvements on firm survival.[9] However, the target set of firms was quite large and potentially included many firms that were inframarginal in terms of survival or employment (Granja and others 2020). This scope suggests that the program may have been poorly sized and targeted, potentially to a large degree. The scale of the program also complicates evaluation of its impacts. If the program was too large and poorly targeted, some standard reduced-form approaches will necessarily compare inframarginal firms with other inframarginal firms, which could miss potentially large impacts on important subsets. Bartlett and Morse (2020) make the case that there is important heterogeneity among PPP target firms and that the social benefits of aiding a large number of very small firms could be large. The program was likely too large, but in our view, the limitations inherent in early evaluations of the PPP's impacts mean it is too early to draw strong conclusions about this magnitude. The potential for heterogeneity among recipient firms, and for longer-term effects on survival and recovery as the pandemic unwinds, means that it is too soon to reject, or accept, large benefits of the program.

I.B. Policies to Provide Financial Support to Households

ECONOMIC IMPACT PAYMENTS Economic Impact Payments (EIPs) were lump-sum payments that aimed to provide broad-based financial relief and economic stimulus that was distributed quickly. Families with income below a threshold ($99,000 for unmarried individuals without children to more than $200,000 for a married couple with children) living in tax filing units where all members had a valid Social Security number (SSN) were eligible for payments. The first payments, issued in April 2020 as part of the CARES Act, provided up to $1,200 per adult and $500 per child under 17. In December 2020, as part of the Consolidated Appropriations

9. Hubbard and Strain (2020) estimate that loan application had economically large impacts on the likelihood of continued operation for smaller businesses over the summer of 2020. Bartik, Cullen, and others (2020) find that PPP receipt increases a firm's assessment of its medium-run survival probability; and cumulative business (and nonbusiness) bankruptcies were slightly lower in 2020 than 2019 (see Administrative Office of the US Courts, Table F-2, "U.S. Bankruptcy Courts—Business and Nonbusiness Cases Commenced, by Chapter of the Bankruptcy Code, during the 12-Month Period Ending December 31, 2019," https://www.uscourts.gov/sites/default/files/bf_f2_1231.2019.pdf, and for the 12-month period ending December 31, 2020, https://www.uscourts.gov/sites/default/files/bf_f2_1231.2020.pdf). Granja and others (2020) find modest effects of PPP disbursement on employment, particularly in late summer 2020, but note that the program likely enhanced firm liquidity, which could promote long-term survival. Chodorow-Reich and others (forthcoming) make a similar point.

Act, 2021, a second payment round provided an additional $600 for each adult and child, with eligibility criteria largely the same as in the first round. While about 70 percent of payments went to households with income less than $100,000, about 30 percent of payments went to families earning more than $100,000.[10]

EIPs were primarily issued through the IRS based on information reported on 2018 and 2019 tax returns. Most taxpayers did not need to request a payment; the IRS issued direct deposits and debit cards automatically. The IRS and Social Security Administration (SSA) partnered to identify retirees and veterans who were eligible but who did not file taxes. Because the payments were distributed through existing systems and based on information that agencies had already collected, benefits were disbursed relatively quickly: 89.5 million payments had been disbursed by the end of April, and 160 million were made by September (US GAO 2020; IRS 2020). In contrast, it took about three months for the stimulus payments in the Great Recession to reach households (US GAO 2008).

Early evidence suggests EIPs helped households maintain their consumption levels. While spending fell across the income distribution between March and April 2020, low-income households' spending rebounded in mid-April, consistent with the timing of the first EIPs (Chetty and others 2020; Cox and others 2020). Households that had an account balance less than $500 spent about 30 percent of their payments within ten days (Baker and others 2020), similar to the spending response for stimulus payments during the Great Recession (Broda and Parker 2014; Parker and others 2013). Household survey responses are consistent with banking data. For example, lower-income Household Pulse Survey respondents are more likely to report spending EIPs, and 80 percent of households that spent the payments purchased essential items, including food, rent, and utilities (Perez-Lopez and Bee 2020). At the same time, the overall personal savings rate sharply increased in the second through fourth quarters of 2020. While higher-income households are more likely to report saving most of the payments (Coibion, Gorodnichenko, and Weber 2020), bank account data also show savings and liquid assets increased across the income distribution after the first EIPs were disbursed (Cox and others 2020).

10. Tax Policy Center, "Combined Effect of Recovery Rebates for Individuals in the COVID-Related Tax Relief Act of 2020 and Coronavirus Aid, Relief, and Economic Security (CARES) Act," table T20-0259, https://www.taxpolicycenter.org/model-estimates/additional-2020-recovery-rebates-individuals-december-2020/t20-0259-combined-effect.

SNAP EMERGENCY ALLOTMENT SNAP is a federal program that aims to prevent hunger and support nutritional intake by providing monthly vouchers to lower-income families that they can use to purchase groceries through a debit card. During the pandemic, policymakers made several reforms to SNAP that expanded eligibility and enabled participants to remain on the program longer. Similar to changes in previous downturns, Congress waived the work requirement for working-age beneficiaries without children in March 2020, allowing these individuals to receive SNAP for more than three months in a three-year period. States were also allowed to extend certification periods and waive interview requirements, which could have increased program retention even among those who remained eligible (Unrath 2021).

Second, the Families First Coronavirus Response Act (FFCRA) included a new expansion, emergency allotment (EA) payments, that allowed states to issue the maximum SNAP amount to all claimants (for their household size) with the additional benefit fully federally funded. The approach of disbursing a single, large benefit was similar in spirit to the uniform UI supplements, with two differences. First, the emergency allotment was voluntary for states (although all opted to participate). Second, unlike the benefit increase during the Great Recession, the emergency allotment did not change benefit amounts for the lowest-income recipients already receiving the maximum benefit.

Both SNAP receipt and benefit amounts increased substantially in 2020, with the number of participating households increasing from 19 million to 23 million and the average benefit increasing from about $238 a month to $349 between September 2019 and 2020. These patterns reflect both the existing program design and reforms made early in the pandemic. First, since SNAP receipt is conditional on income, households that experience income losses become eligible for the program and those already participating may receive higher benefits. Early patterns suggest that SNAP served as an automatic stabilizer during the pandemic recession much like as in previous recessions: as joblessness increased, so did caseloads (Bitler and Hoynes 2016).[11] Second, because the emergency allotment was a change to the existing program, this provision was quickly administered to eligible claimants, with all states issuing EA benefits by mid-April.[12]

11. Estimates of the cyclical increases in SNAP vary widely, but the 2020 increase is generally in line with those (Bitler and Hoynes 2016; Ganong and Liebman 2018).

12. US Department of Agriculture, "USDA Foods: COVID-19 Waivers by State," https://www.fns.usda.gov/disaster/pandemic/covid-19/usda-foods-waivers-flexibilities.

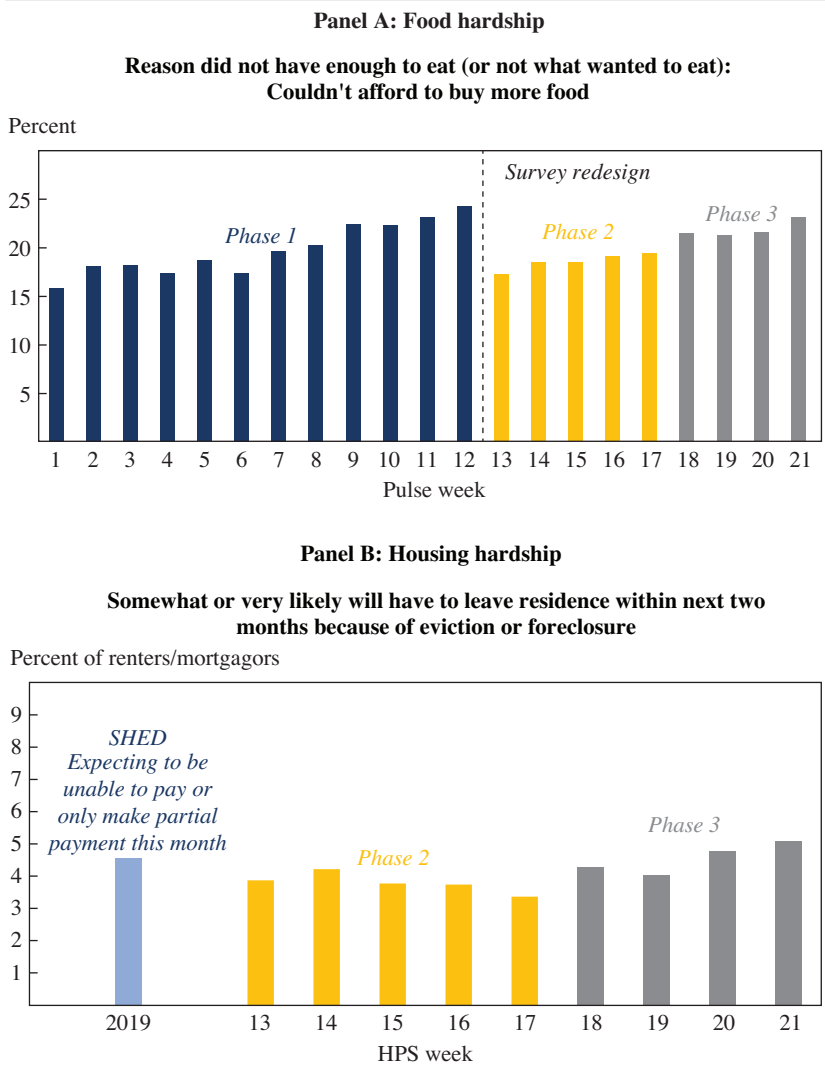
It is difficult to examine in real time whether SNAP expansions reduced food insecurity or helped households meet expenses. However, several patterns suggest that the current amount did not completely address households' nutritional needs. First, although the emergency allotment increased SNAP benefits for some households, the greater benefit amount was partially offset by higher grocery prices (Bitler, Hoynes, and Schanzenbach 2020). In addition, approximately one-third of recipients (those who were already receiving the maximum benefit) did not receive any additional assistance through the emergency allotment (Dean and others 2020). Finally, despite increased SNAP receipt and benefits, food insecurity remained elevated throughout 2020 with more than one in five respondents in the Household Pulse Survey reporting their household experienced food insecurity in the week before the survey. This rate is high compared with prepandemic years, though different approaches to constructing a comparison provide different assessments of the magnitude (Winship and Rachidi 2020). More concerning is the fact that the trend suggests worsening food hardship over the course of the pandemic, shown in figure 2, panel A.

EVICTION MORATORIA High rates of joblessness and income loss prompted concern that the pandemic recession could lead to evictions and foreclosures, putting households at risk for homelessness and housing insecurity. Evictions worsen families' financial situations (Collinson and others 2021; Desmond 2016) and can also lead to homelessness or result in families sharing housing with friends and family members, living "doubled up" (Collinson and Reed 2019). During the pandemic, risk of homelessness became an acute concern for the financial strain it represents but also because congregate living situations—including homeless shelters—place residents and workers at high risk for COVID-19 exposure.

In an effort to prevent evictions and foreclosure, many state and local governments quickly passed moratoria on eviction filings, foreclosures, and utility shutoffs. By April 1, 2020, thirty-nine states and the District of Columbia had passed such relief. By September 4, 2020, when the federal eviction moratorium became effective, forty-five states had passed local bans or moratoria at some point in 2020, twenty-one of which were still effective.[13]

13. E. A. Benfer, R. Koehler, A. K. Alexander, and others, "COVID-19 Eviction Moratoria and Housing Policy: Federal, State, Commonwealth, and Territory" (Google spreadsheet, 2020), https://docs.google.com/spreadsheets/u/1/d/e/2PACX-1vTH8dUIbfnt3X52TrY3dEHQCAm60e5nqo0Rn1rNCf15dPGeXxM9QN9UdxUfEjxwvfTKzbCbZxJMdR7X/pubhtml.

Figure 2. Household Financial Well-Being



Sources: US Census Bureau, Household Pulse Survey; Federal Reserve SHED.

The federal moratorium, delivered through the CDC, established minimum criteria for relief. Renting households earning less than $99,000 (single household, or $198,000 for married couples) that experienced an income loss and would be at risk for homelessness or would be living doubled up with friends and family are covered by the federal rule, but those at risk for foreclosure are not. Under the original rule, households could not be evicted for nonpayment of rent through March 2021; in March 2021, the moratorium was extended through June 2021. When the moratorium expires, all unpaid rent becomes due. State rules can qualify additional households or establish longer relief periods.

Data on households at risk for eviction or foreclosure are limited, both before and during the pandemic. These data limitations, combined with the fact that the federal moratorium and similar state rules are still in effect, make it difficult to observe the full effect of these policies in preventing evictions or maintaining housing stability.[14] With this caveat in mind, about 4 percent of renters or mortgagors believe they are at risk for losing their housing in the near future. Figure 2, panel B, shows this is similar to the share reporting difficulty making their housing payment in 2019, which is a broader group likely to include those fearing eviction. Reed and Divringi (2020) model household budgets and estimate that even with pandemic-era supports, 4 percent of renter households will accrue an average of $5,400 in unpaid rent in 2020, with nonpayment rates higher among single-parent, Black, and Hispanic households. In addition, although utility assistance has increased, utility disconnections and deferred payments sharply increased beginning in mid-2020, indicating households struggle to meet other housing-related expenses (Cicala 2021).

II. Supporting Workers, Families, and Households into 2022

In this section, we identify systematic challenges shared by many of the programs outlined in section I. We provide some general guidance policymakers should consider when developing future stages of policies to facilitate labor market recovery and support household well-being. We focus on

14. Despite the lack of first-stage evidence, some researchers have generated difference-in-differences estimates of the impact of these policies on COVID-19 infection and death rates (Jowers and others 2021). We find the large impacts surprising and in need of better first-stage evidence to support them. Relatedly, such techniques have produced larger impacts of moratoria on deaths than on infections, which seems counterintuitive given the younger populations likely affected.

medium-term actions that are feasible to implement in 2021 and briefly discuss the extent to which recent developments are consistent with these principles. We conclude by discussing some areas that have been overlooked in the response to date and potential responses.

II.A. Payment Infrastructure

We begin with a common success. The programs above, including new ones, disbursed a very large volume of payments quickly, especially relative to historical experience and demonstrating that the United States has considerable ability to rapidly deliver large volumes of support to workers, families, and businesses. The clearest example of this was the EIPs. Within a month of the CARES Act, most eventual beneficiaries had received an EIP, compared to the Great Recession when the first stimulus payments reached households three months after passage (US GAO 2008).

Existing programs provided foundations that allowed for this rapid, large-scale response. For example, SNAP caseloads increased when joblessness rose and more households became eligible. Even new policies were quickly deployed by building on existing programs. The most dramatic example of this was the PPP. By the end of May, more than 70 percent of businesses in the Census Bureau's Small Business Pulse Survey reported receiving PPP loans. This wide reach was possible because the PPP was delivered through existing relationships that essentially enabled the SBA to recruit its network of partner lending institutions to deliver the program, with key support from the Federal Reserve. Expansions to include additional lenders were enacted in subsequent PPP waves to improve reach to the smallest businesses. In the case of SNAP, the USDA also approved most waivers quickly and all states were issuing the new EA payments by mid-April.

Expanded support delivered through the UI system arguably encountered the greatest difficulties. These were the result of known limitations: state-level administration meant that the expansions were unevenly delivered across the United States and outdated computer infrastructure slowed delivery and shaped the assistance that was possible (Botella 2020; McGeehan 2020; Wandner 2018; O'Leary and Wandner 2020). In the case of PUA, fifty-one different agencies had to develop and deploy separate programs in each state, a clear inefficiency. In addition, there were concerns about fraud stemming from cybercrime networks, which stole individuals' identities in order to receive payments. By some estimates, improper payments accounted for about 10 percent of all UI payments

(Clukey 2020).[15] Moreover, many states rely on decades-old technology to administer UI, and federal funding for program administration had fallen over the past decade (Botella 2020). Therefore, additional resources could modernize these systems, improve program integrity, facilitate greater flexibility in program reform, and expedite benefit delivery (Iacurci 2020; McGeehan 2020). As a counterexample, SNAP avoided many technological difficulties and moved all qualifying individuals to the maximum benefit amount without resorting to lump-sum additions to benefits. Despite these shortcomings, UI payments meaningfully increased recipients' spending and their buffered savings almost immediately upon disbursement (Farrell and others 2020, *Consumption Effects*).

II.B. Reaching Marginalized Individuals

Delivering support using existing programs allowed a massive and rapid response but also presented difficulties in reaching individuals and households invisible to existing systems. For example, although administering EIPs through the IRS and SSA allowed a timely disbursement to most households, these agencies could not automatically identify eligible households not captured by either system—primarily the estimated 12 million nonveteran, working-age individuals with income below the amount required to file federal income taxes (Marr and others 2020). Bhutta and others (2020) show that the 2020 programs were highly effective at restoring financial security for working families but barely improved security for households with no working adults. Although the IRS created a non-filers tool that collected individuals' SSNs and mailing information to determine eligibility in an attempt to identify this population, data from agencies that administer other safety net programs, primarily SNAP and Medicaid, could identify other eligible individuals (Augustine, Davis, and Ramesh 2021; Marr and others 2020).

A straightforward way to improve reach to marginalized populations is to use Medicaid enrollment information. Compared to other income assistance programs, Medicaid serves families farther up the income distribution, allowing policymakers to identify households that may qualify for EIPs or SNAP that are not participating in other programs. In addition, there is a precedent for using this information; for example, some states have

15. See also Thompson Reuters, "Unemployment Fraud in the Future Phases of COVID-19," https://legal.thomsonreuters.com/en/insights/articles/phases-of-unemployment-after-covid.

successfully used administrative Medicaid data to establish school meal eligibility.[16] The Census Bureau, which has now matched census records to IRS filings, is another source of information on those who might be eligible for payments but invisible in IRS data. In many cases, addresses and information on non-filing individuals could be retrieved from these merged data.

A related challenge stemmed from issuing checks and debit cards to individuals who did not have bank account information on file with the IRS or SSA. While these households still received EIPs, they received payments several weeks later than those with direct deposit information. This delay could be shortened by expanding the non-filer tool to allow individuals to register an e-payment platform on which to receive their benefit (Cook 2020).

In the case of UI, PUA expanded UI eligibility to populations that were not previously eligible. This is a considerable expansion of coverage, accounting for approximately 40 percent of all continuing claims as of January 2021 (Bureau of Labor Statistics 2021). However, take-up is likely incomplete and for some populations, language barriers or inadequate access to technology could have presented additional barriers to take-up, both in UI and in other programs.[17] For example, in California, UI claims would have been 23 percent higher if the UI recipiency rate was the same across the state as it was in wealthy neighborhoods (Bell and others 2020). In other cases, program reforms left out some of the most vulnerable households. For example, the lowest-income SNAP recipients did not benefit from the emergency allotment in 2020, and unemployed workers with the lowest earnings were ineligible for additional benefits through LWA. From a macroeconomic perspective, leaving out the most vulnerable households could have ramifications for the broader economic recovery, as these households have the highest marginal propensity to consume.

II.C. Targeting Infrastructure

The flip side of the rapid distribution of large dollar amounts in support was limited targeting across almost all programs. For example, about 30 percent of EIPs went to households with incomes above $100,000 who

16. USDA, "Evaluation of the Direct Certification with Medicaid for Free and Reduced-Price Meals (DCM-F/RP) Demonstrations, Year 2," https://www.fns.usda.gov/nslp/evaluation-direct-certification-medicaid-free-and-reduced-price-meals-dcm-frp

17. The high rates of efficacy in Bhutta and others (2020)—defined as providing sufficient liquid resources that enable households to weather a sustained income loss, given the existing distribution of emergency savings—required that households can access the full set of benefits for which they qualify. Limited access, most likely in UI, could reduce these rates.

mostly saved, rather than spent, the payments (Coibion, Gorodnichenko, and Weber 2020). On the firm side, PPP access was granted to relatively large firms that could likely have weathered the pandemic without a forgivable loan. The stated rationale for a lack of targeting was timely delivery. Even when aid was delivered through existing programs, the need to register or apply, as with UI and the PPP, slowed delivery compared to programs where aid was disbursed automatically based on administrative data, as with the EIPs. However, targeting need not depend on lengthy or difficult information to verify applications. The EIP program, particularly in later waves, could have improved targeting using available administrative data, and the PPP could have improved targeting by simply narrowing the participation criteria. In the longer term, improved application and program IT infrastructure could reduce this time to delivery by enhancing opportunities for administrative targeting.[18] Administration support is important to the long-term success of UI, as recent infrastructure challenges showed.[19] It may also be important to future aid disbursements through the Treasury, as the IRS has experienced a long period of underinvestment.

In the next twelve months, policymakers should consider how to combine rapid delivery with identifying groups where aid will be the most impactful. Part of the rationale for modest targeting was to improve receipt among marginalized individuals and households. But reports that significant shares of households still face food and shelter insecurity suggest that the broad approach is not adequately supporting households with the least resources. This is unlikely to be solved by continued broad disbursements. Legislation enacted in early 2021 lowered the income threshold for EIPs, taking one step toward more targeting. However, several options for getting support to those with greatest need were not discussed. One option is to leverage administrative data to target support according to dimensions other than annual (2019) income. Ideally, policymakers should define eligibility on characteristics that are observable, not subject to manipulation, and correlated with earnings capacity or losses in the pandemic. For example,

18. Delivery infrastructure improvements that speed registration and verification could also address cybersecurity issues in support delivery. Some states have reported concerns about significant numbers of fraudulent UI claims, potentially facilitated by hackers using information from previous large consumer data breaches (Cohen 2020).

19. It is possible that the low take-up of short-time compensation could have resulted in part from the surge in the regular UI program that overwhelmed state UI offices given infrastructure limitations. For instance, the expansions in the regular program pushed many offices to their capacity and might have left little scope for promoting the existing STC program. Alternatively, generous PPP aid might have crowded out firm demand for STC.

in the near term, policymakers could target programs to households who were working in certain industries prior to the pandemic, living in communities most affected by the pandemic, or who experienced large drops in income between 2019 and 2020. Data on all of these indicators are available in the same tax and Social Security information that inform the current EIP program.

II.D. Designing Phaseouts and Automatic Stabilizers

It is possible that expanding support to the economy will prove easier than withdrawing assistance. Policymakers should therefore begin considering how best to phase out programs and expansions as the economy recovers.

So far, most policy lapses have occurred at arbitrary dates established by the original legislation, rather than when the labor market or economy reaches a certain milestone. This pattern has resulted in "cliffs" where recipients incur sharp reductions in benefits after a certain day and potentially huge income uncertainty. Introducing automatic stabilizers that peg program changes to the state of the economy can avoid these cliffs and improve confidence in the economy among households using support programs.

The pandemic economy poses some additional challenges to traditional proposals for automatic stabilizers, which often focus on adjusting unemployment insurance or other benefits as unemployment rates fluctuate. Labor force participation rates have fallen steeply in the pandemic. As workers who withdraw from the labor force are not included in the unemployment rate, this statistic may be a limited indicator of economic recovery in the current setting. For this reason, policymakers might consider a less conventional approach to automatic winding down that incorporates the nature of the pandemic recession. For example, a set of withdrawal phases could incorporate a combination of improvements in employment rates, decreases in unemployment rates, and declines in virus prevalence. Extensions enacted in the March 2021 American Rescue Plan (ARP) largely retain fixed program end dates.

Another consideration is whether programs should phase out at the same time or in a particular order. In our view, staggered withdrawal of these programs is preferred to expirations that occur simultaneously because the current income support provisions provide few work disincentives but distortions from the large UI expansions and unprecedented PPP support are likely larger. As noted earlier, evidence shows that these distortions have not prolonged the initial recovery, but as the economy continues to improve, these programs might lead to greater departures from normal economic

activity. On this basis, the PPP should be rolled back first, followed by various UI expansions, with some UI reforms, like the PEUC supplements, ending before others, like PUA and the extended benefits duration. An intermediate step might provide partial UI payments to those unable to return to their previous hours or earnings levels. Other household supports that are less tied to work should be reformed last, though direct cash support could be increasingly targeted to households and individuals with greater needs. This phased approach should encourage labor market reentry and business expansion while continuing household support to bolster consumer demand and provide an income backstop. The ARP partially adopts this order. It essentially reissues the 2020 EIPs, provides for UI supplements that are about half those in 2020, and reduces PPP funding to a small fraction of the 2020 level. On the other hand, the supplemental amounts expire at the same time as other UI extensions, creating a bundled program cliff, and expirations are not tied to the health of the labor market, creating risk that the withdrawal of support could be either too fast or too slow. Both are aspects of the 2021 policy response that could be further improved.

Phasedown considerations are also paramount in the context of the federal eviction moratorium, although there is considerable uncertainty about the number and extent of housing disruptions it has prevented. Patterns from expiration of local moratoria indicate the number of eviction filings increase after filing bans lapse, suggesting eviction rates may increase once the federal moratorium is lifted (Cowin, Martin, and Stevens 2020). This policy has not received as much attention as other supports, likely since it comes at no direct cost to taxpayers. However, it is potentially an important component of households' balance sheets, and its phaseout should be considered alongside reductions in other out-of-work supports. The ARP provides $25 billion in rental assistance, but it is unclear whether or how this will replace moratoria supports. Moreover, since housing relief is largely administered at a local level, the effectiveness of these funds in maintaining housing stability will vary with local infrastructure quality.

III. Conclusion and Additional Steps

The support delivered to US workers and households during the pandemic has been historic and has prevented dire hardship for millions. In spite of some notable successes, this response has overlooked some key needs and features of the pandemic.

First, policies to date have done little to develop systems for weathering and containing coronavirus outbreaks as the broader recovery progresses.

The course of COVID-19 infections is still unclear, but the presence of variants means outbreaks may continue for some time. This is an opportune time to assess how best to encourage firms, local governments, and other enterprises to invest in systems to allow for containment or even early detection of future outbreaks. At a minimum, firms should be supported in offering paid leave for employees who need to receive testing or isolate. State and local governments could be encouraged to adopt plans for future temporary local shutdowns as a condition of receiving federal aid.

The pandemic will have lasting effects, particularly through persistent effects of COVID-19 and through the negative impacts of a prolonged period of poor mental health. Medium-run policymaking should consider how to expand existing programs to meet these needs. Those with persistent COVID-19-related health deficits could be covered through expansions to temporary disability insurance programs until more is known about the course of these effects. Strategies for expanding access to mental health care should be explored—particularly for parents, whose mental health may affect outcomes for their children. Those who have suffered severe trauma in the pandemic—such as those who lost friends and family members, served in hard-hit hospitals, or experienced the year's spike in homicides—are other groups for whom ongoing services should be considered. Expanding existing systems could help meet some of this additional need. For example, the second group might be well served by expanding access to mental health care through the US Department of Veterans Affairs or other trauma specialists. Enhanced mental health care for parents might be provided through schools.

In addition, it is likely that the extended schooling disruptions will have lasting impacts on skills for millions of children. This is a key time to consider interventions to reverse this deficit. Funding could be allocated to schools to offer remedial services, perhaps through expanded summer programs. The ARP provides some funding for this, and states and districts would benefit from guidance on how best to use it.

Finally, given the scope and scale of the pandemic response, it is critical we continue to evaluate these efforts to understand the full extent of their reach, which populations were helped, who was left out, and how local responses shaped the successes and shortcomings. To this end, the medium-term responses should build in data transparency requirements. This need is particularly great regarding UI recipients and beneficiaries of the eviction moratoria, since even basic data on receipt are limited in these cases.

New and existing data collection and sharing by federal agencies have allowed researchers to piece together a picture of the state of the economy

and financial well-being relatively quickly. Our assessment is that the 2020 social insurance system response had many successes, but there are several feasible adjustments that could improve the reach and efficiency of these programs in 2021. It is also time for policymakers to consider when and how to roll back these programs, in order to give Americans a sense of the path back to normal economic activity.

ACKNOWLEDGMENTS The authors thank Katharine Abraham, Phillip Swagel, Zach Swaziek, David Cho, Winnie van Dijk, and participants at the spring *BPEA* conference for helpful input. All remaining errors are the authors'.

References

Altonji, Joseph, Zara Contractor, Lucas Finamor, Ryan Haygood, Ilse Lindenlaub, Costas Meghir, Cormac O'Dea, Dana Scott, Liana Wang, and Ebonya Washington. 2020. "Employment Effects of Unemployment Insurance Generosity during the Pandemic." Working Paper. https://tobin.yale.edu/sites/default/files/files/C-19%20Articles/CARES-UI_identification_vF(1).pdf.

Augustine, Elsa, Charles Davis, and Aparna Ramesh. 2021. "The Stimulus Gap: 2.2 Million Californians Could Miss $5.7 Billion in Federal Stimulus Payments." Policy Brief. Los Angeles: California Policy Lab. https://www.capolicylab.org/wp-content/uploads/2021/04/The-Stimulus-Gap-in-California.pdf.

Autor, David, David Cho, Leland Crane, Mita Goldar, Byron Lutz, Joshua Montes, William Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz. 2020. "An Evaluation of the Paycheck Protection Program Using Administrative Payroll Microdata." Working Paper. https://economics.mit.edu/files/20094.

Baker, Scott R., Robert A. Farrokhnia, Steffen Meyer, Michaela Pagel, and Constantine Yannelis. 2020. "Income, Liquidity, and the Consumption Response to the 2020 Economic Stimulus Payments." Working Paper 27097. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27097.

Bartik, Alexander W., Marianne Bertrand, Feng Lin, Jesse Rothstein, and Matt Unrath. 2020. "Measuring the Labor Market at the Onset of the COVID-19 Crisis." *Brookings Papers on Economic Activity*, Summer, 239–326.

Bartik, Alexander W., Zoe B. Cullen, Edward L. Glaeser, Michael Luca, Christopher T. Stanton, and Adi Sunderam. 2020. "The Targeting and Impact of Paycheck Protection Program Loans to Small Businesses." Working Paper 27623. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27623.

Bartlett, Robert P., III, and Adair Morse. 2020. "Small Business Survival Capabilities and Policy Effectiveness: Evidence from Oakland." Working Paper 27629. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27629.

Bell, Alex, Thomas J. Hedin, Geoffrey Schnorr, and Till von Wachter. 2020. *An Analysis of Unemployment Insurance Claims in California during the COVID-19 Pandemic*. Los Angeles: California Policy Lab. https://www.capolicylab.org/wp-content/uploads/2021/02/Dec-21st-Analysis-of-CA-UI-Claims-during-the-COVID-19-Pandemic.pdf.

Bhutta, Neil, Jacqueline Blair, Lisa Dettling, and Kevin Moore. 2020. "COVID-19, the CARES Act, and Families' Financial Security." *National Tax Journal* 73, no. 3: 645–72.

Bitler, Marianne, and Hilary Hoynes. 2016. "The More Things Change, the More They Stay the Same? The Safety Net and Poverty in the Great Recession." *Journal of Labor Economics* 34, no. S1, pt. 2: S403–44.

Bitler, Marianne P., Hilary W. Hoynes, and Diane Whitmore Schanzenbach. 2020. "The Social Safety Net in the Wake of COVID-19." *Brookings Papers on Economic Activity*, Summer, 119–45.

Boone, Christopher, Arindrajit Dube, Lucas Goodman, and Ethan Kaplan. 2021. "Unemployment Insurance Generosity and Aggregate Employment." *American Economic Journal: Economic Policy* 13, no. 2: 58–99.

Botella, Elena. 2020. "Why New Jersey's Unemployment Insurance System Uses a 60-Year-Old Programming Language." *Slate*, April 9. https://slate.com/technology/2020/04/new-jersey-unemployment-cobol-coronavirus.html.

Broda, Christian, and Jonathan A. Parker. 2014. "The Economic Stimulus Payments of 2008 and the Aggregate Demand for Consumption." *Journal of Monetary Economics* 68:S20–S36.

Bureau of Labor Statistics. 2021. "Unemployment Insurance Weekly Claims." News Release. February 4. https://oui.doleta.gov/press/2021/020421.pdf.

Chen, Haiqiang, Wenlan Qian, and Qiang Wen. 2021. "The Impact of the COVID-19 Pandemic on Consumption: Learning from High Frequency Transaction Data." *American Economic Association Papers and Proceedings* 111:307–11.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Michael Stepner, and the Opportunity Insights Team. 2020. "The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data." Working Paper 27431. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27431.

Chodorow-Reich, Gabriel, Olivier Darmouni, Stephan Luck, and Matthew C. Plosser. Forthcoming. "Bank Liquidity Provision across the Firm Size Distribution." *Journal of Financial Economics*. https://scholar.harvard.edu/chodorow-reich/publications.

Cicala, Steve. 2021. "The Incidence of Extreme Economic Stress: Evidence from Utility Disconnections." Working Paper 28422. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28422.

Clukey, Keshia. 2020. "Covid-19 Scammers Driving Unemployment Fraud across Nation." *Bloomberg Law*, October 19. https://news.bloomberglaw.com/daily-labor-report/covid-19-scammers-driving-unemployment-fraud-across-nation.

Cohen, Patricia. 2020. "Fraud Schemes Exploit Weak Spots in Unemployment Claims System." *New York Times*, October 1. https://www.nytimes.com/2020/10/01/business/economy/unemployment-benefits-fraud.html.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2020. "How Did US Consumers Use Their Stimulus Payments?" Working Paper 27693. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27693.

Collinson, Robert, John Eric Humphries, Nicholas Mader, Davin Reed, Daniel Tannenbaum, and Winnie van Dijk. 2021. "Eviction and Poverty in American Cities: Evidence from Chicago and New York." Working Paper.

Collinson, Robert, and Davin Reed. 2019. "The Effects of Evictions on Low-Income Households." Working Paper. https://robcollinson.github.io/RobWebsite/jmp_rcollinson.pdf.

Cook, Lisa. 2020. "Getting Money Urgently to Low-Wage U.S. Workers." Blog post, March 30, Washington Center for Equitable Growth. https://equitablegrowth.org/getting-money-urgently-to-low-wage-u-s-workers/.

Cortes, Guido Matias, and Eliza Forsythe. 2020. "Impacts of the COVID-19 Pandemic and the CARES Act on Earnings and Inequality." Working Paper 13643. Bonn: IZA Institute of Labor Economics. http://ftp.iza.org/dp13643.pdf.

Council of Economic Advisers. 2014. *The Economic Impact of the American Recovery and Reinvestment Act Five Years Later*. https://obamawhitehouse.archives.gov/sites/default/files/docs/erp_2014_chapter_3.pdf.

Couture, Victor, Jonathan I. Dingel, Allison Green, Jessie Handbury, and Kevin R. Williams. 2020. "Measuring Movement and Social Contact with Smartphone Data: A Real-Time Application to COVID-19." Working Paper 35. Federal Reserve Bank of Minneapolis, Opportunity and Inclusive Growth Institute. https://www.minneapolisfed.org/research/institute-working-papers/measuring-movement-and-social-contact-with-smartphone-data-a-real-time-application-to-covid-19.

Cowin, Rebecca, Hal Martin, and Clare Stevens. 2020. "Measuring Evictions during the COVID-19 Crisis." Federal Reserve Bank of Cleveland. www.clevelandfed.org/en/newsroom-and-events/publications/community-development-briefs/db-20200717-measuring-evictions-during-the-covid-19-crisis.aspx.

Cox, Natalie, Peter Ganong, Pascal Noel, Joseph Vavra, Arlene Wong, Diana Farrell, Fiona Greig, and Erica Deadman. 2020. "Initial Impacts of the Pandemic on Consumer Behavior: Evidence from Linked Income, Spending, and Savings Data." *Brookings Papers on Economic Activity*, Summer, 35–69.

Davies, Paul S., and William R. Morton. 2020. "COVID-19 and Direct Payments to Individuals: Economic Impact Payments (EIPs) for Social Security and Supplemental Security Income Beneficiaries." Washington: Congressional Research Service. https://crsreports.congress.gov/product/pdf/IN/IN11290.

Dean, Stacy, Lauren Hall, Brynne Keith-Jennings, and Dottie Rosenbaum. 2020. *SNAP Benefit Boost Would Get Needed Aid to the Poorest Participants, Who Have Been Left Out*. Washington: Center on Budget and Policy Priorities. https://www.cbpp.org/research/food-assistance/snap-benefit-boost-would-get-needed-food-aid-to-the-poorest-participants.

Desmond, Matthew. 2016. *Evicted: Poverty and Profit in the American City*. New York: Crown.

Elgin, Cyehun, Gokce Basbug, and Abdullah Yalaman. 2020. "Economic Policy Responses to a Pandemic: Developing the COVID-19 Economic Stimulus Index." *Covid Economics: Vetted and Real-Time Papers* 3 (April 10): 40–54.

Fairlie, Robert W., and Frank Fossen. 2021. "Did the $660 Billion Paycheck Protection Program and $220 Billion Economic Injury Disaster Loan Program Get Disbursed to Minority Communities in the Early Stages of COVID-19?" Working Paper 28321. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28321.

Farrell, Diana, Peter Ganong, Fiona Greig, Max Liebeskind, Pascal Noel, Daniel Sullivan, and Joseph Vavra. 2020. "The Unemployment Benefit Boost: Initial Trends in Spending and Saving When the $600 Supplement Ended." Policy Brief. Washington: JPMorgan Chase Institute. https://www.jpmorganchase.com/

content/dam/jpmc/jpmorgan-chase-and-co/institute/pdf/Institute-UI-Benefits-Boost-Policy-Brief_ADA.pdf.

Farrell, Diana, Peter Ganong, Fiona Greig, Max Liebeskind, Pascal Noel, and Joseph Vavra. 2020. *Consumption Effects of Unemployment Insurance during the COVID-19 Pandemic*. Washington: JPMorgan Chase Institute. https://www.jpmorganchase.com/institute/research/labor-markets/unemployment-insurance-covid19-pandemic.

Ganong, Peter, and Jeffrey B. Liebman. 2018. "The Decline, Rebound, and Further Rise in SNAP Enrollment: Disentangling Business Cycle Fluctuations and Policy Changes." *American Economic Journal: Economic Policy* 10, no. 4: 153–76.

Ganong, Peter, Pascal Noel, and Joseph Vavra. 2020. "US Unemployment Insurance Replacement Rates during the Pandemic." *Journal of Public Economics* 191:104273.

Giupponi, Giulia, and Camille Landais. 2020. "Subsidizing Labor Hoarding in Recessions: The Employment and Welfare Effects of Short-time Work." Discussion Paper 1585. London: Centre for Economic Performance. https://cep.lse.ac.uk/pubs/download/dp1585.pdf.

Goolsbee, Austan, and Chad Syverson. 2021. "Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020." *Journal of Public Economics* 193:104311.

Granja, João, Christos Makridis, Constantine Yannelis, and Eric Zwick. 2020. "Did the Paycheck Protection Program Hit the Target?" Working Paper 27095. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27095.

Hedin, Thomas J., Geoffrey Schnorr, and Till von Wachter. 2020 "An Analysis of Unemployment Insurance in California during the COVID-19 Pandemic." Policy Brief. Los Angeles: California Policy Lab. https://www.capolicylab.org/wp-content/uploads/2020/06/June-11th-Analysis-of-CA-UI-Claims-During-the-COVID-19-Pandemic.pdf.

Hubbard, Glenn, and Michael R. Strain. 2020. "Has the Paycheck Protection Program Succeeded?" *Brookings Papers on Economic Activity*, Fall, 335–78.

Iacurci, Greg. 2020. "Unemployed Workers Could Get More than 100% of Their Paycheck under the Coronavirus Bill." CNBC, March 27. https://www.cnbc.com/2020/03/27/how-unemployed-workers-could-get-more-than-100percent-of-their-paycheck.html.

Internal Revenue Service (IRS). 2020. "Treasury, IRS Deliver 89.5 Million Economic Impact Payments in First Three Weeks, Release State-by-State Economic Impact Payment Figures." News Release, April 28. https://www.irs.gov/newsroom/treasury-irs-deliver-89-point-5-million-economic-impact-payments-in-first-three-weeks-release-state-by-state-economic-impact-payment-figures.

Jowers, Kay, Christopher Timmins, Nrupen Bhavsar, Qihui Hu, and Julia Marshall. 2021. "Housing Precarity and the Covid-19 Pandemic: Impacts of Utility Disconnection and Eviction Moratoria on Infections and Deaths across US

Counties." Working Paper 28394. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28394.

Krolikowski, Pawel, and Anna Weixel. 2020. *Short-Time Compensation: An Alternative to Layoffs during COVID-19*. Federal Reserve Bank of Cleveland. https://www.clevelandfed.org/ec202026.

Liu, Haoyang, and Desi Volker. 2020a. "Where Have the Paycheck Protection Loans Gone So Far?" Blog post, May 6, Liberty Street Economics, Federal Reserve Bank of New York. https://libertystreeteconomics.newyorkfed.org/2020/05/where-have-the-paycheck-protection-loans-gone-so-far.html.

Liu, Haoyang, and Desi Volker. 2020b. "The Paycheck Protection Program Liquidity Facility (PPPLF)." Blog post, May 20, Liberty Street Economics, Federal Reserve Bank of New York. https://libertystreeteconomics.newyorkfed.org/2020/05/the-paycheck-protection-program-liquidity-facility-ppplf.html.

Marr, Chuck, Kris Cox, Kathleen Bryant, Stacy Dean, Roxy Caines, and Arloc Sherman. 2020. *Aggressive State Outreach Can Help Reach the 12 Million Non-Filers Eligible for Stimulus Payments*. Washington: Center on Budget and Policy Priorities. https://www.cbpp.org/research/federal-tax/aggressive-state-outreach-can-help-reach-the-12-million-non-filers-eligible.

McGeehan, Patrick. 2020. "He Needs Jobless Benefits. He Was Told to Find a Fax Machine." *New York Times*, April 4. https://www.nytimes.com/2020/04/04/nyregion/coronavirus-ny-unemployment-benefits.html.

Moffitt, Robert A., and James P. Ziliak. 2020. "COVID-19 and the US Safety Net." *Fiscal Studies* 41, no. 3: 515–48.

O'Leary, Christopher J., and Stephen A. Wandner. 2020. "An Illustrated Case for Unemployment Insurance Reform." Working Paper 19–317. Kalamazoo, Mich.: Upjohn Institute. https://research.upjohn.org/cgi/viewcontent.cgi?article=1336&context=up_workingpapers.

Parker, Jonathan A., Nicholas S. Souleles, David S. Johnson, and Robert McClelland. 2013. "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review* 103, no. 6: 2530–53.

Perez-Lopez, Daniel, and Charles Adam Bee. 2020. "Majority Who Received Stimulus Payments Spending Most of It on Household Expenses." America Counts, June 24. https://www.census.gov/library/stories/2020/06/how-are-americans-using-their-stimulus-payments.html.

Reed, Davin, and Eileen Divringi. 2020. *Household Rental Debt during COVID-19*. Federal Reserve Bank of Philadelphia. https://www.philadelphiafed.org/community-development/housing-and-neighborhoods/household-rental-debt-during-covid-19.

Strain, Michael R. 2020. "The Bad and the Good in the New Payroll Protection Program." *Bloomberg Opinion*, December 21. https://www.bloomberg.com/opinion/articles/2020-12-21/payroll-protection-program-costly-errors-welcome-corrections.

Unrath, Matthew. 2021. "Targeting, Screening, and Retention: Evidence from California's Food Stamps Program." Working Paper. https://mattunrath.github.io/files/research/Unrath_SNAP.pdf.

US Government Accountability Office (GAO). 2008. *Tax Administration: IRS's 2008 Filing Season Generally Successful Despite Challenges, although IRS Could Expand Enforcement during Returns Processing*. Washington. https://www.gao.gov/assets/gao-09-146.pdf.

US Government Accountability Office (GAO). 2020. *Financial Audit: IRS's FY 2020 and FY 2019 Financial Statements*. Washington. https://www.gao.gov/assets/gao-21-162.pdf.

Wandner, Stephen A., ed. 2018. *Unemployment Insurance Reform: Fixing a Broken System*. Kalamazoo, Mich.: W. E. Upjohn Institute for Employment Research.

Winship, Scott, and Angela Rachidi. 2020. *Has Hunger Swelled?* Washington: American Enterprise Institute. https://www.aei.org/research-products/report/has-hunger-swelled/.

GENERAL DISCUSSION With respect to Andrew Atkeson's model, Robert Hall noted another scenario which ought to be included: delays in the propagation of the vaccine. There are two phases to testing a vaccine, Hall continued: the safety phase—which is quite straightforward as it doesn't require a large sample and can be done on an ongoing basis—and the efficacy phase—which, in the case of the United States, was delayed. Indeed, Hall observed, the vaccine was created in the first two months of 2020, right at the beginning of COVID-19. Thus, Hall concluded, delays in vaccine propagation were in the testing process, which suggests that a counterfactual with faster propagation would likely make a large difference in Atkeson's model.

Hall then wondered about Christina Romer's question regarding the prospects in the expansion of the economy. On the labor market side, Hall reflected that with around 6 percent current unemployment, it is possible to get back to an unemployment rate of about 4 percent, suggesting 2 percentage points of expansion. However, he pointed out that during the pandemic there has been a large decrease in the size of the labor force, implying that as the pandemic recedes there will be an increase in the participation rate. Hall specified that this is not usually the case during a typical expansion.

Claudia Sahm considered a point that Romer made during her discussion about the efficacy of the individual stimulus checks. Sahm specified that typically when there is a demand shock—which Sahm argued was the case during the past year—the best thing to do was very stimulative fiscal policy, such as stimulus checks. Furthermore, Sahm observed, the US safety net doesn't reach a lot of unemployed workers—as Abigail Wozniak also mentioned in her discussion—and she thus wondered why stimulus checks weren't the best policy choice as they do seem to reach unemployed workers falling through the cracks.

Katharine Abraham pondered the limited use of short-time compensation. In her view, given the advantages to reducing workers' hours and giving them prorated benefits rather than laying workers off, this limited use was a lost policy opportunity. The advantages of short-time compensation include allowing workers to avoid the costs of long-term unemployment and helping businesses maintain relationships with their employees, making it easier for them to ramp back up when the economy recovers. Especially as the businesses that were most affected by the pandemic provide services that will still be needed in the future, Abraham stated that she wasn't especially worried about the use of short-time compensation slowing needed reallocation. She asked Wozniak and Romer to comment on whether making

greater use of short-time compensation as a response to temporary downturns would be good future policy.

Addressing Atkeson, Maurice Obstfeld observed that one can become even more pessimistic when one understands that variants are likely to be endogenous and can arise even when a large proportion of the population is infected or immunized. He brought up the case of the city of Manaus in Brazil where the Brazilian variant arose even when an estimated two-thirds or more of the population had been infected. Obstfeld reasoned that there are two pockets of people that are of special concern currently: the vaccine-hesitant and those living in developing countries. Obstfeld remarked that the Kaiser Family Foundation estimated in March that developing countries have only been able to purchase enough doses to vaccinate about a third of their populations, suggesting other variants might emerge.[1]

Steven Davis commented on the "pipes" of the unemployment insurance (UI) benefits system, addressing a recent report from the Inspector General of the Department of Labor.[2] This report, Davis stated, contains two findings relevant to the panel. First, the report finds that in fourteen out of the last seventeen years the improper payment rate of the UI system has been 10 percent or more. Second, the report states that their initial audit under the Coronavirus Aid, Relief, and Economic Security Act suggests this number might be higher. Connecting these two findings with Romer's comments on the size of UI benefit payments in 2020, Davis reasoned, suggests that the government misdirected something on the order of $90 billion or more through the UI system. All in all, Davis expressed, there is a huge problem in functioning of the UI system, which ultimately undermines the targeting goal of the UI benefit program and involves a large waste of funds.

Frederic Mishkin addressed Atkeson, remarking that his model did not distinguish between different mitigation measures although some types of mitigation are much less costly than other types. For example, Mishkin expressed, mask mandates may be cheap and relatively effective, while lockdowns are quite costly. Similarly, Mishkin stated, testing and tracing is more expensive than mask mandates. Mishkin asked Atkeson to comment

1. Anna Rouw, Adam Wexler, Jennifer Kates, and Josh Michaud, "Global COVID-19 Vaccine Access: A Snapshot of Inequality," Kaiser Family Foundation, March 17, https://www.kff.org/policy-watch/global-covid-19-vaccine-access-snapshot-of-inequality/.

2. Department of Labor, Office of the Inspector General, "DOL-OIG Oversight of the Unemployment Insurance Program," https://www.oig.dol.gov/doloiguioversightwork.htm, accessed March 21, 2021.

on these three types of mitigation measures: mask mandates, testing and tracing, and lockdowns.

Gary Burtless asked Wozniak whether there is any evidence on trends in charitable giving, especially to organizations—like food banks or soup kitchens—that address immediate spending needs. After all, Burtless concluded, many people and richly endowed institutions have seen sizeable increases in wealth since January 2020.[3]

David Wilcox wondered whether a better approach to massive fiscal support is to introduce a temporary broad-based program of partial income insurance, up to some capped amount.

Joshua Gotbaum commented on modifications of the Paycheck Protection Program (PPP) that relaxed the requirement for firms to maintain their payrolls. Gotbaum observed that different analyses have reached varying conclusions about the PPP's efficacy in maintaining employment, and he inquired about Wozniak's view.

Erica Groshen addressed Wozniak and Romer, remarking that targeting is more possible with good granular and timely information, as well as robust, agile program infrastructure. Groshen speculated whether another lesson from the pandemic is that official statistics and programmatic operational infrastructure—such as the UI system—have a high value during crises. This added value, she deduced, is evidence for adequate investment in them on an ongoing basis as a means of increasing resilience to shocks.

3. After the conference, Burtless noted that charitable gifts—in general and those aimed specifically at victims of the COVID-19 pandemic—did rise in 2020 despite the sharp downturn in the economy, according to the latest data provided by the Lilly Family School of Philanthropy at Indiana University; Anna Pruitt and John Bergdoll, "Americans Gave a Record $471 Billion to Charity in 2020, amid Concerns about the Coronavirus Pandemic, Job Losses and Racial Justice," blog post, Lilly Family School of Philanthropy, June 16, 2021, https://blog.philanthropy.iupui.edu/2021/06/16/americans-gave-a-record-471-billion-to-charity-in-2020-amid-concerns-about-the-coronavirus-pandemic-job-losses-and-racial-justice/.

ANDREW METRICK
Yale University

DANIEL TARULLO
Harvard University

Congruent Financial Regulation

ABSTRACT After the global financial crisis, bank regulation became more stringent, and as a result the traditional banking system was well capitalized leading into the COVID-19 pandemic. But these same regulatory changes also incentivized a continuing migration of traditional banking activities to nonbank financial institutions (NBFIs), where looser regulation allowed for dangerous buildups of systemic risk. These risks were then realized across many NBFIs and markets in 2020. While legislation to harmonize regulation across these different domains would be desirable, we do not believe it likely in the foreseeable future. In this paper we propose a congruence principle for financial regulation, whereby regulators use existing statutory authority to coordinate rules across economically similar instruments. We provide examples of how such congruence could work for the cases of nonprime mortgage finance and the markets for US Treasury securities.

The prominent role of nonbank financial institutions (NBFIs) in the global money market panic produced by the first COVID-19 lockdown in March 2020 reprised the part they played in the global financial crisis of 2007–2009. Collectively, NBFIs and their associated short-term wholesale funding markets constitute a large and growing component of the global financial system (Financial Stability Board 2020b). But regulation has not kept up with this growth. Howell Jackson (1999) observed more than twenty years ago that the regulatory constraints applicable to financial intermediation are more a function of the classification of the institution within which the intermediation is conducted than its fundamental nature

Conflict of Interest Disclosure: The authors did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this paper. They are currently not officers, directors, or board members of any organization with an interest in this paper. Discussant Hyun Song Shin is an economic adviser and head of research at the Bank for International Settlements, which had the right to review his work.

and risks. That observation remains apposite today. Indeed, as the two intervening financial disruptions have shown, the consequences of a patchwork regulatory system have become substantially more serious. While more rigorous prudential standards implemented after the global financial crisis made banks a source of stability in the spring of 2020, the vulnerabilities of NBFIs were such that the Federal Reserve felt it had no choice but to use its emergency powers to create an astonishing range of market-supporting measures (Federal Reserve Board 2020a).

The freezing of so many financing markets in March 2020 has revived interest in NBFI activities and calls for action across key parts of the official sector (Bank of England 2020; Federal Reserve Board 2020b; Financial Stability Board 2020b; Financial Stability Oversight Council 2020; International Monetary Fund 2020). This renewed attention, while belated in some instances, suggests enough consensus to sustain momentum for a regulatory response. Our argument here is that, to be effective, regulatory initiatives cannot replicate the largely reactive and ad hoc approach to NBFI activities followed after the global financial crisis. Intermediation activities can quickly migrate in response to regulatory change. The astonishing range of market-supporting measures adopted by the Federal Reserve and other central banks in the first half of 2020 has demonstrated that in very bad states of the world these activities will be supported ex post, regardless of whether they have been regulated ex ante. Hence the need for a proactive regulatory approach that can, in Jeremy Stein's memorable phrase in a different context, get "in all of the cracks" (Stein 2013).

In this paper we urge a cohesive approach to the macroprudential regulation of NBFI activities. Specifically, we propose that an overarching congruence principle should inform, and unify, regulatory efforts to address the contribution of NBFIs to systemic risk. Under this principle, the regulation of economically similar activities would be coordinated across agencies, with the goals of minimizing regulatory arbitrage and ensuring that the social costs of systemic risk are internalized by private actors, regardless of their institutional form.

We note at the outset that we are not opining in this paper on the appropriate stringency of financial regulation motivated by financial stability concerns. It is of course the case that the capital regime after the global financial crisis has made banks safer (International Monetary Fund 2018). A full year into the pandemic, large US banks remain well capitalized and, thus far, free of runs or any short-term concerns about solvency. In our view, this experience is a good place to begin in thinking about the degree of resiliency the regulatory system should mandate. But here we address

a different issue. We argue that the regulatory structure is incomplete, and that incompleteness—that *incongruence*—leads to risk shifting and the creation of alternative pathways of financial intermediation that are neither planned nor optimal. The regulatory reform project is unfinished, and that project needs some organizing principles.

Section I lays out the main elements of the congruence principle. In the past, capital and other prudential requirements were justified primarily as microprudential tools, the logic being that the government needed to protect its implicit and explicit insurance position for individual institutions. Under that logic, there was no argument for imposing congruent regulations on nonbank activities, since those institutions were outside of the explicit safety net. But the events of this century have demonstrated the macroprudential value of capital standards. Our proposed principle would be applied through imposing regulation based on the substantive nature of the intermediation. Its application would be limited to regulation motivated by the containment of systemic risk. It calls for regulation to be *congruent*, not necessarily *identical*. This is an important distinction: congruent regulation makes use of economically similar (but not identical) instruments, with regulation coordinated across agencies. We believe that congruence is both more flexible and more achievable than other alternatives.[1]

Section II provides two case studies of the role played by important NBFIs in recent events. It explains how the lack of regulatory congruence contributed to both the buildup of risk and market dysfunction during stress. The first case study is the evolution of nonprime mortgage finance in recent decades, a classic example of the disintermediation of banks from their core function of deposit taking and lending. The second case study—Treasury securities—looks at a similar process of bank disintermediation from banks' function as primary dealers of government securities. In both cases, the markets changed over the preceding ten years, such that the pathways for financial intermediation changed radically. Both case studies illustrate how market-driven capital levels were insufficient to reassure markets during the COVID-19 crisis and how a liquidity crisis was quelled only after extraordinary action by the Federal Reserve under its emergency liquidity powers.[2]

1. Identical rules would be the default under various forms of activity-based regulation. Our proposal for congruent regulation could also be categorized as instrument-based regulation, as in Acharya and Öncü (2012).

2. See Barth and Kahn (2020), Fleming and Ruela (2020), and Haddad, Moreira, and Muir (2020). The notion that liquidity spirals can occur even for instruments with no underlying risk was introduced by Morris and Shin (2008).

Section III provides an example of how the congruence principle could be implemented in what we believe would be a significant first step—an alignment of bank capital requirements with the rules for margining at clearinghouses and haircuts for repo transactions. Our proposed approach to margining shows how similar contributions to systemic risk could be addressed through regulations applicable to all financial intermediaries regardless of legal form, chartering identity, or business model. In this exercise, we take current regulatory structure and statutory authority as fixed. From that starting place, we provide a road map for the regulatory actions that would be necessary to achieve congruence. To execute on this road map, we must meet various legal and institutional challenges. In section IV, we conclude the paper with a few observations on what is at stake. If congruence cannot be achieved under the current configuration of administrative agency authorities, then either we must pass new laws or we must accept that systemically risky NBFI activities will continue to evolve well ahead of our balkanized regulatory system. A glossary (in the online appendix) provides definitions for key terms and acronyms.

I. A Congruence Principle for Financial Regulation

The congruence principle is a starting point for a regulatory response to the secular trend of financial intermediation migrating outside the banking system. This development has reproduced the same risks of rapid reduction of system-wide liquidity, asset fire sales, and adverse impacts on other intermediaries that motivate macroprudential regulation of banks. Indeed, the very strengthening of bank regulation to contain these risks has increased opportunities for arbitrage (Barth and Kahn 2020; Financial Stability Board 2020a). The absence of an effective regulatory response will reinforce risk-taking tendencies across these markets and potentially erode the franchise value of the regulated institutions whose risk taking has been constrained so as to limit negative externalities and moral hazard.

The congruence principle can be stated simply: forms of financial intermediation posing similar risks to financial stability should be regulated with similar stringency, regardless of legal form, chartering identity, or business model. The amount of systemic risk contributed by nonbank intermediation should be contained to levels reflecting the same balance between shorter-term growth and medium-term financial stability considerations that is implicitly incorporated in prudential regulation of banking organizations.

Our conception of this principle is that it (1) applies only to systemic risk; (2) requires congruent but not necessarily identical modes of regulation;

and (3) operates in an ex ante, comprehensive fashion. In this paper we apply the principle only to the discrete, though important, area of lending collateralized with securities. However, we believe congruence would be a useful informing principle in determining appropriate regulatory responses across the range of NBFI activities, including the form of maturity transformation unaccompanied by risk transformation that is characteristic of money market and other mutual funds.

The focus on systemic risk has two implications for implementing the principle. First, not all nonbank intermediation would be subject to prudential regulation—only those forms that pose enough risk to warrant the costs involved in devising and applying a regulatory framework. Second, congruence measures would be derived only from those elements of prudential banking regulation directed at reducing runs, fire sales, and contagion more generally. Banks are still special in numerous respects. They benefit from federally insured deposit insurance and provide transaction accounts to most households and businesses. The failure of very large banks and the holding companies of which they are part would give rise to financial instability. Thus congruence measures would decidedly not aim to replicate for nonbanks the entire range of bank regulations.

The aim of congruent, but not necessarily identical, regulation arises from the focus on systemic risk, but has somewhat broader implications. For example, we would probably not apply a form of the liquidity coverage ratio (LCR) or similar bank requirements to money market funds (Li and others 2020; COVID-19 Market Impact Working Group 2020). The LCR helps prevent excessive dependence on short-term funding and fire sales, to be sure. But it serves other purposes as well, such as providing breathing space to government authorities deciding how to deal with a highly stressed bank. The COVID-19 crisis confirmed the view of many critics that gates, fees, and maturity limits on assets are inadequate responses to the funding vulnerabilities, and consequent contribution to systemic risk, of money market mutual funds. But other modes of regulation more fitting to the risks of the money fund business model could achieve results congruent to the systemic risk protection afforded by the LCR.

Similarly, in our discussion of congruence for Treasury-backed securities lending later in this paper, we do not propose identical regulatory measures. Capital requirements, margining, and haircuts can all serve the purpose of inhibiting procyclical excesses and ensuring resiliency of financial firms. Yet each is not equally suited to universal application, even for similar risks. Notably, capital regulation is viable only where the market actor engaging in a form of risk-carrying transaction is subject to regulation

on a firm-wide basis. Moreover, the risks associated with, say, a Treasury-backed repo vary with such contextual circumstances as whether the transaction is part of a netting arrangement. In short, while we share the starting point of many other commentators concerned that financial regulation has been based more on form than on function, we doubt that an entirely function-based approach would be feasible.

The congruence principle should operate in an ex ante, comprehensive fashion so as to capture evolutions of funding practices, intermediary structures, and other financial innovations before they grow into problems. Following the global financial crisis, regulators generally considered each form of NBFI activity in relative isolation. Given the relative ease with which funding can be redeployed to new investment vehicles, this approach almost guarantees that regulators will be several steps behind emerging risks. Implementation of the congruence principle should aim to address maturity and risk transformation in sufficiently broad terms that financial innovations contributing to systemic risk will presumptively be covered. The architects of those innovations would of course be welcome to ask regulators for modifications of regulations tailored to the details of the new form or practice or to argue that no systemic risk is created. But by making the default situation one in which regulation applies, this attribute of the congruence principle would provide a timelier check on regulatory arbitrage and the accretion of systemic risk.

Each of these attributes of the congruence principle will entail both policy judgment and practical hurdles. Judgment will be required in determining, for example, which forms of intermediation contribute materially to systemic risk and whether a congruent but not equivalent form of regulation achieves a roughly equivalent reduction in systemic risk. More generally, it is unrealistic to think that a truly comprehensive framework can be devised and then implemented from the outset. Indeed, our experience in policymaking inclines us to believe that such a complex effort would bog down from its inception and, even if it could be achieved institutionally, would almost surely produce a bevy of unintended, undesirable consequences.

With these reasons for caution in mind, we regard the attributes of the congruence principle as more aspirational than immediately achievable, especially the attribute of comprehensiveness. But the principle is both a good starting point and a good lodestar for building out and regularly adjusting the regulation of systemic risk. We now turn to the two case studies—nonprime mortgage lending and Treasury security markets—and then to a discussion of how congruence could be achieved in these specific markets.

II. Case Studies

II.A. Nonprime Mortgage Lending

In this case study, we look at the evolution of a part of the mortgage lending business, and how regulation influenced that evolution. Our example focuses on the nonprime component of the market, where the loans are ineligible for guarantees from Freddie Mac or Fannie Mae. We consider three methods of intermediation for such mortgages. We follow other authors in ascribing some of the impetus for development of the second method to regulatory arbitrage and believe that the third method can similarly be explained in part as a response to regulation after the global financial crisis.

As shown in figure 1, we begin with the most straightforward method of bank finance, which was dominant until the end of the twentieth century. In this figure—and similar ones to follow—we use ovals to denote players in the intermediation chain, rectangles to denote regulators, number labels on exchanges of cash or securities, and letter labels for regulatory relationships.

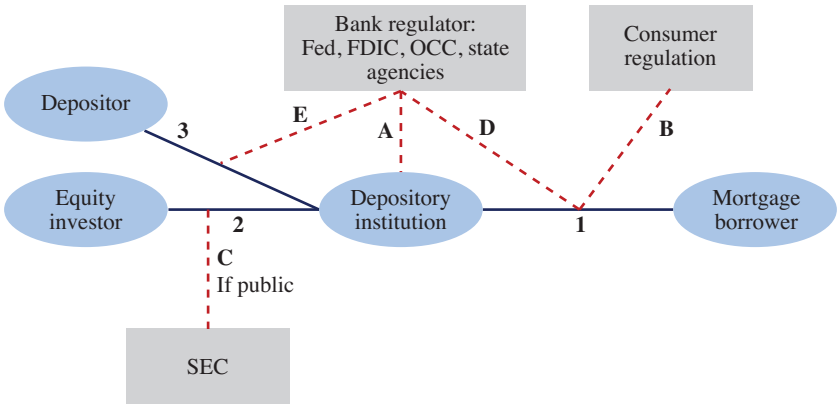
Here, a bank makes mortgage loans (step 1) financed by some combination of equity (step 2) and deposits (step 3). The mortgages then stay on the bank's balance sheet. To regulate this activity, bank regulators set capital requirements as a function of the characteristics of specific mortgages (step A).[3] Other regulation would come from the consumer side (step B, both federal and state) and from the Securities and Exchange Commission (SEC; step C) if the bank's equity is publicly traded.[4] This list of players and regulatory relationships is not intended to be exhaustive; instead, we want to highlight some key features that allow useful comparisons with different ways to perform the same economic function. For our purposes here, the most important regulation is that the bank would be required to fund some part of each mortgage with its own equity, at a ratio that has varied over time but has always been strictly positive for nonprime loans.

Figure 2 illustrates the version of this intermediation that captured so much market share in the years prior to the global financial crisis: private-label securitization funded by highly rated debt securities. Here, we have replaced the bank in the center of the figure with a generically named mortgage company, which is not itself a depository institution. It may be either

3. See 78 Fed. Reg. 62018, 78 Fed. Reg. 55340, 79 Fed. Reg. 20754, and Government Accountability Office (2016).

4. See Dodd-Frank Act Pub. L. No. 111-203, Title X (2010), NY Banking Law § 10 (2012), and Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338, 1385-1407 (1999).

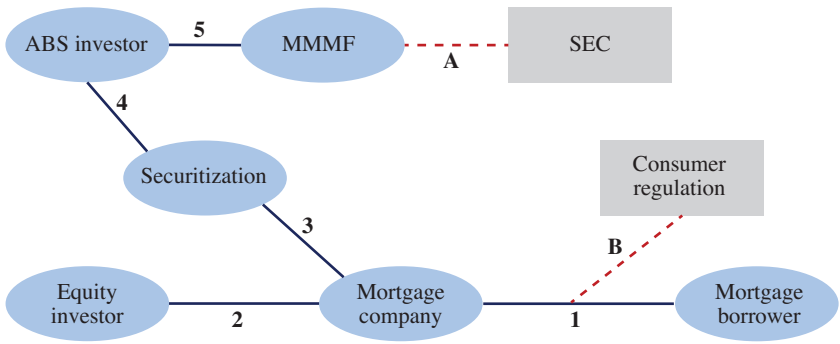
Figure 1. Traditional Mortgage Lending



Source: Authors.

Note: A depository institution originates a mortgage to the borrower in step 1, and this is regulated by a consumer regulatory agency and the appropriate bank regulator (B and D). The depository institution funds the mortgage with a combination of deposits and equity (2 and 3). Other regulation of this mortgage lending process comes from the SEC over the equity investment (C) and the bank regulator (A and E).

Figure 2. Mortgage Lending before the Global Financial Crisis



Source: Authors.

Note: A mortgage company originates the mortgage to the borrower in step 1. This transaction is subject to consumer regulations (B). The mortgage company funds its lending through a combination of equity (2) and the securitization of the mortgage loans (3). These securities are sold to asset-backed securities investors (4) and money market mutual funds (5), which are regulated by the SEC (A).

unrelated to a bank holding company (BHC) or an unconsolidated affiliate not subject to capital regulation. In either case, there are no regulatory capital requirements placed on the funding of the mortgage company. In this form of intermediation the mortgage itself will never sit on the balance sheet of the mortgage company but instead will be transferred to a securitization trust, shown in step 3.[5] The trust assets are then divided into layers of seniority, with the vast majority being highly rated debt sold to asset-backed security investors in step 4. The securitized bonds received by those investors have an average maturity similar to the mortgages that underlie them, but these investors often perform their own maturity transformation by issuing short-term debt to money market mutual funds (MMMFs) in step 5. Overall, figure 2 represents the intermediation chain that grew rapidly before the global financial crisis and crashed terribly as that crisis began. All along this chain, the only capital requirements are those imposed by the market: the borrowers in step 1 might make only a minimal (or zero) down payment, the mortgage company is not subject to any capital regulation, and in practice the securitization trust will need to satisfy only the rating agencies. Figure 3 shows the rise and fall of several of these links in the chain: private-label securitization, asset-backed commercial paper, and MMMFs (both prime and government).

The precrisis developments in mortgage finance have been well studied, but the postcrisis shift in nonprime mortgage finance has received far less scholarly attention.[6] The scale and scope of these changes have been remarkable. Figure 4 illustrates the main pathway in this market as of March 2020.

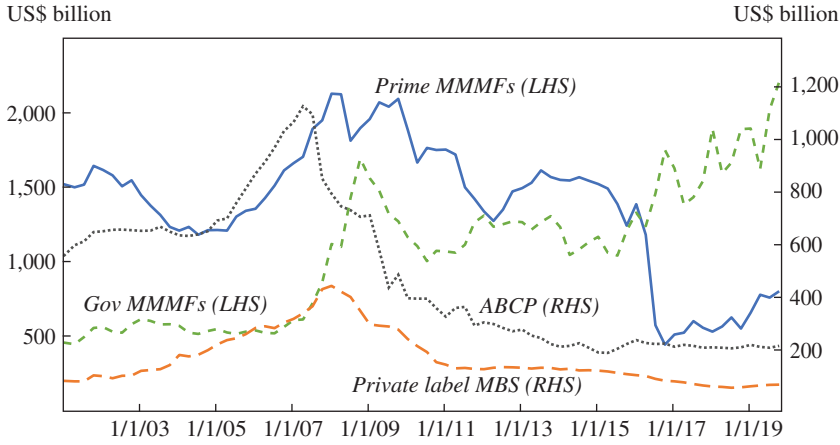
Figure 4 reflects several key changes from the chain shown in figure 2. First, the nonprime borrowing now occurs mostly through loans guaranteed by the Federal Housing Administration (FHA) or the Department of Veterans Affairs (VA) and then securitized by Government National Mortgage Association, or Ginnie Mae (GNMA). This change is illustrated in figure 5.

Second, bank-affiliated mortgage companies were another casualty of the global financial crisis, as postcrisis regulatory reforms both tightened requirements for consolidation and increased required capital levels

5. In this example, the securitization trust actually takes on all of the risk, and the underwriter is then out of the chain. But some of the transfers done before the global financial crisis carried various forms of implicit guarantees and made these relationships more complex. See Acharya, Schnabl, and Suarez (2013).

6. Kim and others (2018) and Gete and Reher (2021) are notable exceptions, and their work is closely related to the pathway described in figure 4.

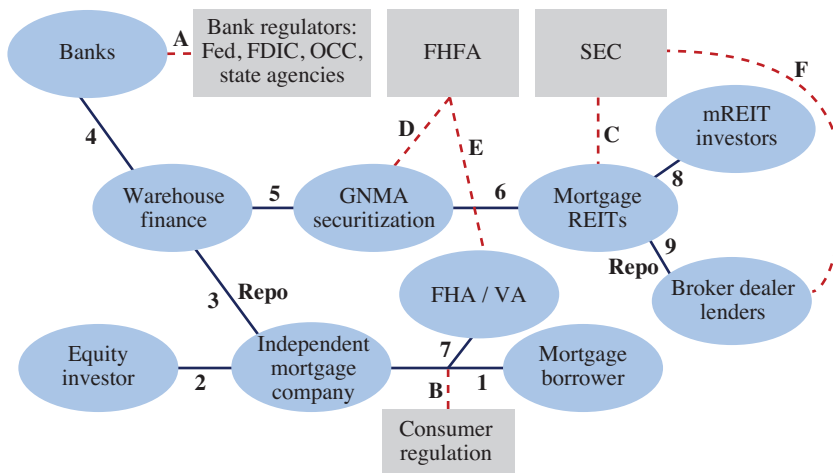
Figure 3. Prime and Government MMMFs, Asset-Backed Commercial Paper, and Private-Label Mortgage-Backed Securities



Sources: Federal Reserve; iMoneyNet.

Note: Prime money market mutual funds (MMMFs) invest in non-Treasury assets, such as commercial paper issued by corporations or agency securities. Government MMMFs invest at least 99.5 percent of securities in cash, government securities, or repo agreements collateralized by government securities or cash. Asset-backed commercial paper is a short-term investment vehicle that is a form of commercial paper collateralized by other financial assets. Private-label mortgage-backed securities are securitized mortgages that do not conform to criteria defined by the government-sponsored enterprises.

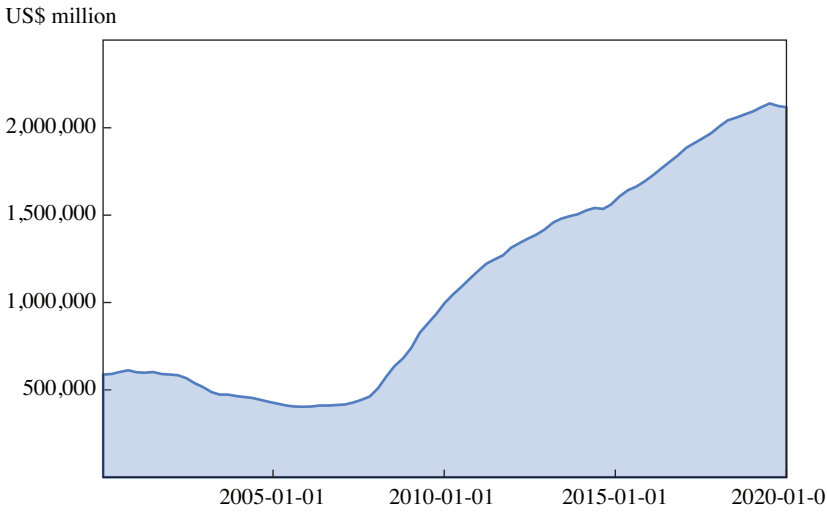
Figure 4. Mortgage Lending after the Global Financial Crisis



Source: Authors.

Note: The independent mortgage company makes a mortgage to the borrower (1) which is subject to consumer regulation (B), and these loans are guaranteed by the FHA/VA. The mortgage company funds the mortgage with a combination of equity (2) and repo (3). The mortgages are securitized by Ginnie Mae (5), and a large portion are purchased by mortgage real estate investment trusts (6), which are regulated by the SEC (C).

Figure 5. Total Mortgages Held in a Ginnie Mae Pool



Source: Federal Reserve.

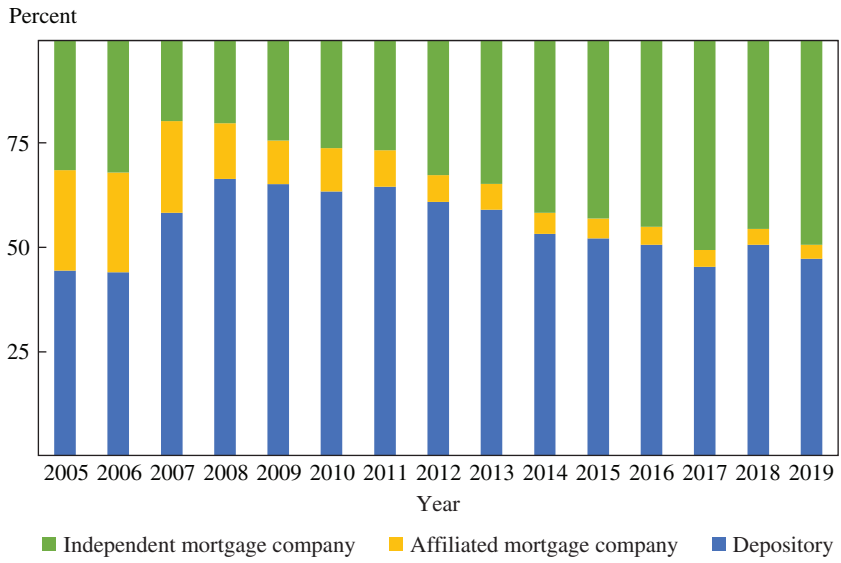
Note: Ginnie Mae, formally known as the Government National Mortgage Association, is a self-financing, government-sponsored enterprise that guarantees investors the timely payment of principal and interest on securities backed by federally insured or guaranteed loans (primarily loans insured by the Federal Housing Administration or guaranteed by the Department of Veterans Affairs). The data show the current asset level of mortgages in Ginnie Mae's pool—mortgages are added, and mortgages roll off as they are paid off.

(Kim and others 2018). The majority of these nonprime loans are now arranged by independent mortgage companies. This change is illustrated in figure 6.

Finally, a large portion of the GNMA securitizations is now purchased by mortgage real-estate investment trusts (mREITs), often highly leveraged and financed mostly with short-term repo contracts. This specialized investment trust barely existed in 2000 before growing to almost $300 billion in assets prior to the global financial crisis. Unlike other mortgage players from that era, the mREIT industry bounced back after the crisis, partly on the back of friendly regulatory treatment for the rehypothecation for GNMA securities, their main source of investment.[7] Panel A of figure 7 shows the mREIT industry reaching nearly $700 billion prior to the pandemic. These mREIT assets are low-yielding, often government-guaranteed securities. The high absolute returns earned to drive growth were generated

7. Gete and Reher (2021) discuss the importance of this regulatory change for the VA/FHA loans that underlie GNMA securitizations.

Figure 6. Share of Dollar Amount of Loan Originations by Lender Type



Sources: Home Mortgage Disclosure Act (HMDA); Robert Avery; and authors' calculations.

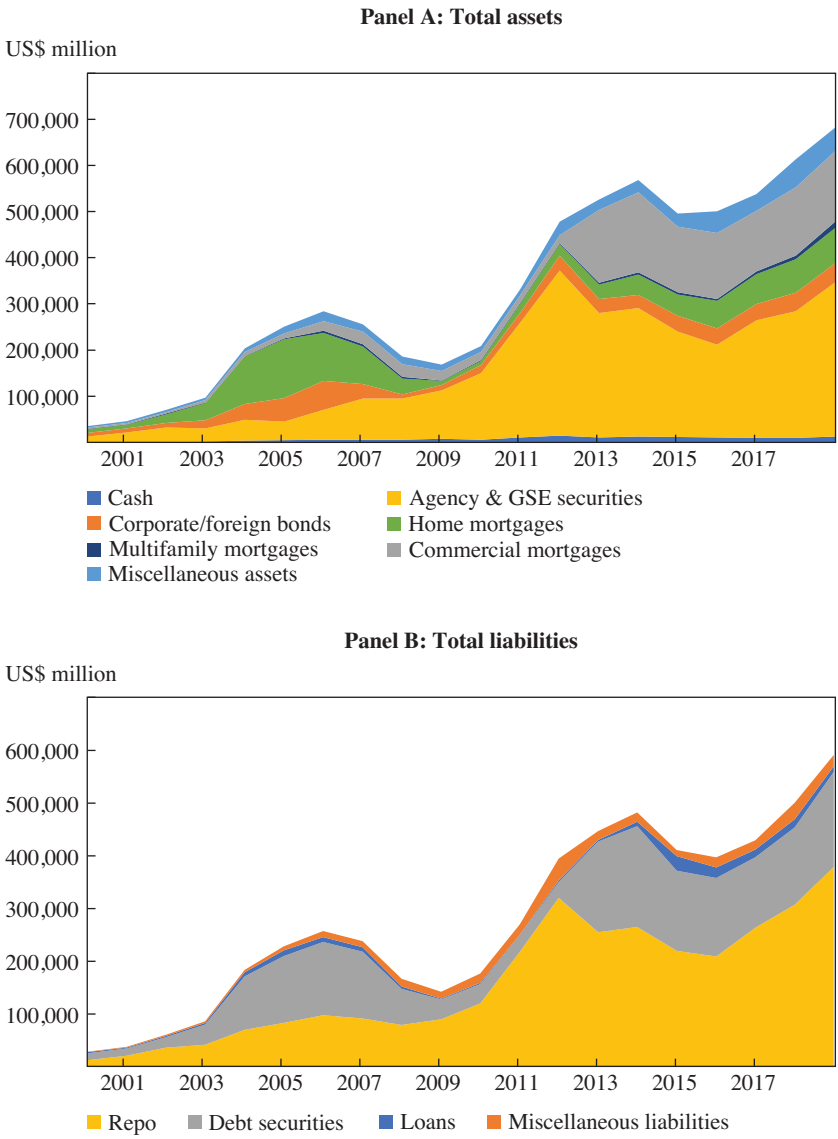
Note: A variety of lending financial institutions can originate mortgages. An independent mortgage company is an institution that typically does not take deposits from the public and focuses its business on mortgage lending. An affiliated mortgage company is an institution that typically does not take deposits from the public and focuses its business on mortgage lending but is owned by or affiliated with a bank or credit union. A depository institution is an institution that takes deposits from the public and originates mortgages.

primarily by leverage. This leverage is unregulated and provided primarily by repo, as shown in panel B.[8] In March 2020, we learned again that highly leveraged institutions, even with safe assets on their balance sheet, can be a casualty of a generalized panic.

Figure 8 illustrates the turmoil for mREITs in March 2020, when agency-focused mREITs lost about 80 percent of their value. Since these vehicles invest only in government-guaranteed instruments, these extreme losses are driven solely by liquidity problems. In the commotion, some mREITs had their collateral seized by lenders, and many others would have suffered the same fate had the panic not been stopped by the Federal Reserve's massive injection of liquidity into unregulated parts of the financial system (Scaggs 2020; Maloney 2020; Hoffman and Zuckerman 2020).

8. For more details on mREIT structure and strategy, see Pellerin, Sabol, and Walter (2013).

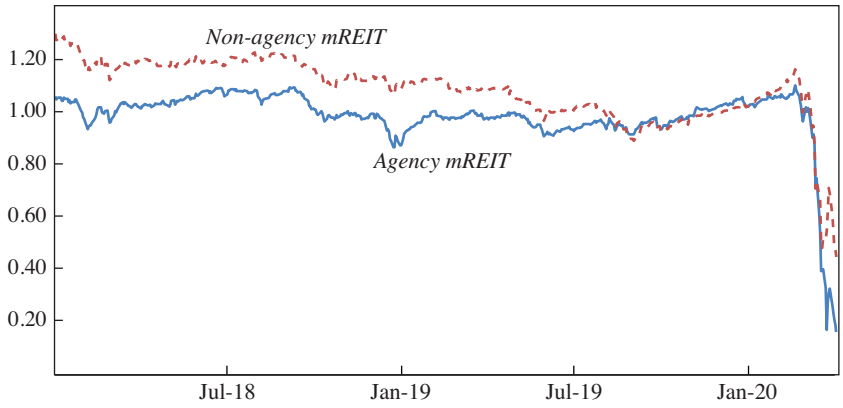
Figure 7. Mortgage Real Estate Investment Trusts (mREITs)



Source: Federal Reserve.

Note: A mortgage real estate investment trust (mREIT) invests in mortgages and mortgage-backed securities to earn income from the investments. An mREIT holds the mortgages and MBS on its balance sheet and funds its investments with debt, repo, or equity capital.

Figure 8. Mortgage Real Estate Investment Trust Equity Prices



Source: Bloomberg.

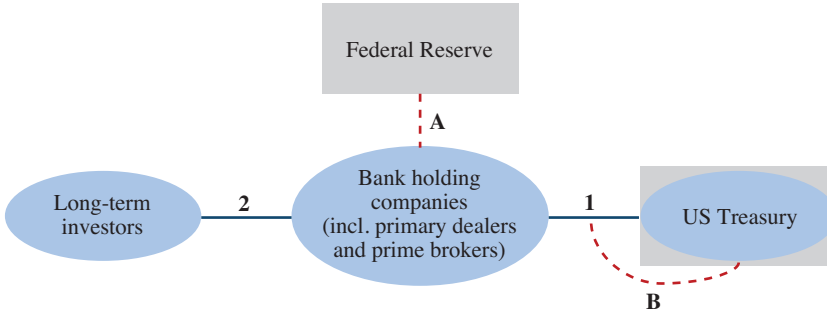
Note: Agency mREITs invest in mortgages and mortgage-backed securities backed by one of the government-sponsored enterprises (such as Fannie Mae or Freddie Mac). Non-agency mREITs invest in mortgages and mortgage-backed securities that are not backed by a government guarantee.

II.B. The Treasury Market in March 2020

In this case study, we examine the impact of incongruent regulation on the market for US Treasury securities. Here, our focus is on banks' role as agents: the broker-dealers that facilitate the distribution of Treasury securities between the US government and the ultimate investor.

Prior to the global financial crisis, most government securities were intermediated by primary dealers, the largest of which were affiliates of BHCs. Following the global financial crisis, enhanced prudential standards, higher capital requirements, and changes in banks' own risk management policies placed pressure on this role at the same time that Treasury debt was rapidly increasing, leaving a gap to be filled by NBFIs (Duffie 2018, 2020; Financial Stability Oversight Council 2020; Klingler and Sundaresan 2020). In this case, it was hedge funds that stepped in, making markets in Treasuries through a multistep chain of exchange-traded futures contracts and repo-financed long positions in physical Treasuries (Barth and Kahn 2020, 2021; Financial Stability Board 2020a; Kothari and others 2020; Schrimpf, Shin, and Sushko 2020). The BHCs remained in the chain through their prime-broker subsidiaries, with a complex and shifting impact on their capital requirements. Then, in just the past few years, a portion of this repo activity moved to a central clearinghouse and completely away from bank balance sheets. As we all learned in March 2020, even the market for the world's safest security can malfunction during a stress event. In this

Figure 9. Dealer-Bank Intermediation of Government Securities



Source: Authors.

Note: The US Treasury sells bonds through auctions to primary dealers—the bank holding companies (1). The Treasury also has regulatory authority of the market (B). Primary dealers sell most of the Treasury securities to long-term investors (2). Bank holding companies are subject to capital regulation from the Federal Reserve (A).

instance, the malfunction elicited a targeted and overwhelming response from the Federal Reserve, which calmed the market through measures that (directly and indirectly) rescued many NBFI players (Financial Stability Oversight Council 2020; Financial Stability Board 2020a; Federal Reserve Board 2020c; Kothari and others 2020).

Less than a year after the event, scholars and practitioners have generated a large literature that assesses the underlying causes of the breakdown and proposes policy solutions to prevent a recurrence (Duffie 2020; Liang and Parkinson 2020). We are not engaging here in this important debate over optimal changes in the Treasury market. There have been many excellent suggestions, but our purpose is different: to use this event as a salient example of incongruent regulation, demonstrating that we stumbled into a system of intermediation that was both fragile and unplanned. To do this, we first sketch three different pathways for the process of Treasury debt intermediation.

Figure 9 is a schematic of the players and regulatory relationships in the Treasury market in its simplest form. Here, Treasury sells bonds through auctions (step 1), with primary dealers as the main buyers, who then ultimately sell most of the securities to long-term investors (step 2). The largest primary dealers are subsidiaries of bank holding companies and as such are subject to capital regulation by the Federal Reserve (A). Note the dual role here of the Treasury, which is both the seller of securities and a regulator of those security markets. The importance of this dual role is highlighted below.

The "peace dividend" of the 1990s reduced federal deficits and Treasury issuance, which was part of the reason for the growth of securitization and the manufacture of safe-asset substitutes in the years leading up to the global financial crisis (Gorton and Metrick 2010, 2012). In recent years, the imbalance has gone the other way, with sharply increasing issuance of Treasuries in the years leading up to the COVID-19 crisis (Liang and Parkinson 2020). This additional issuance required ever growing balance sheet capacity from the primary dealers, as they needed to hold ever larger inventories between auctions and the eventual sales**.**

For the simple intermediation shown in figure 9, regulatory balance sheet constraints were nonbinding prior to the global financial crisis. The capital requirements binding on large banks were risk-adjusted measures of assets, and the risk adjustment on Treasuries was not material.[9] Following the global financial crisis and the implementation of the Basel III accords in the United States, large banks became subject to a supplementary leverage ratio, where the computed leverage encompassed all assets, including reserves and Treasury securities.[10] Banks of global systemic importance are subject to an enhanced supplementary leverage ratio (ESLR) of 5 percent at the holding company level and 6 percent in bank affiliates. In the years immediately following implementation of the ESLR, it was often considered the binding constraint on balance sheet space, effectively implying that a bank subject to the ESLR would need to hold 5 percent capital against any inventory of Treasury securities (Quarles 2018).[11] To the extent that banks consider such capital to be costly, this provided an incentive for the shift to and growth of alternative pathways for the intermediation of Treasuries.

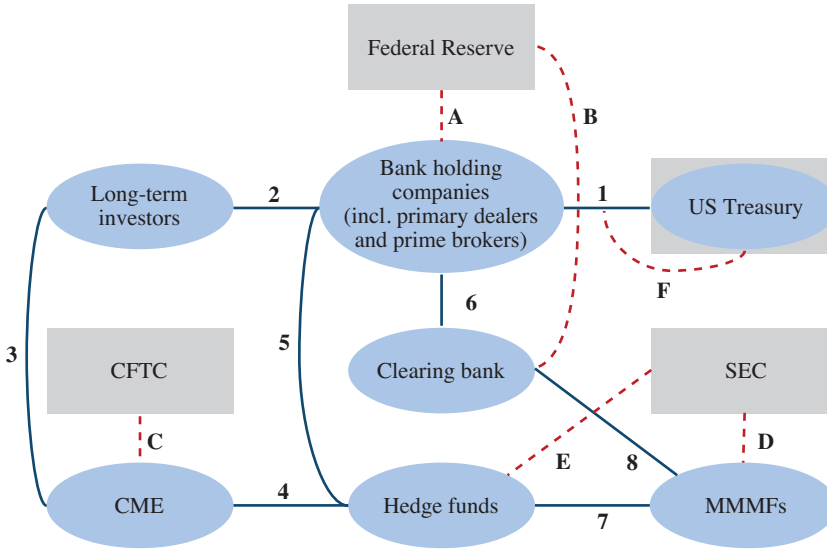
Figure 10 illustrates one popular alternative, where investors looking for the most liquid Treasury market turned more toward the futures contracts traded on the Chicago Mercantile Exchange (CME), shown here in step 2. This increased liquidity raised the price of Treasury futures relative to their cash equivalent and introduced an arbitrage basis between the two. The gap left by primary dealers is taken up (in part) by relative-value hedge funds,

9. Treasuries held in the banking book are zero weighted, that is, considered free of credit risk. Treasuries held in the trading book, especially those of longer maturity, do have a positive market risk weighting, but it is quite small.

10. Regulatory Capital Rule 79 FR 24528.

11. Similarly, until the Federal Reserve eliminated the requirement that banks meet a minimum leverage ratio requirement as part of stress test–related capital requirements, the post-stress leverage ratio was often the binding capital constraint for between two and four of the global systematically important banks.

Figure 10. Nonbank Intermediation of Government Securities



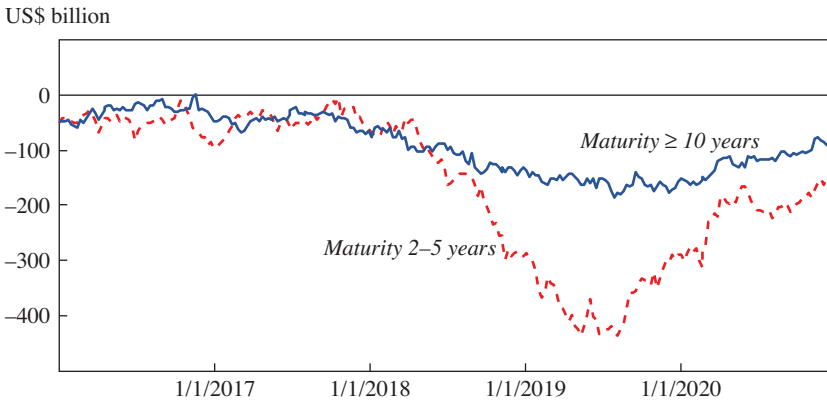
Source: Authors.

Note: The US Treasury sells bonds through auctions to primary dealers—the bank holding companies (1). The Treasury also has regulatory authority of the market (F). In this diagram, hedge funds short the Treasury futures contract and purchase Treasury securities. Hedge funds finance this "relative value" trade through bilateral repos with their prime brokers. The prime broker will often rehypothecate the Treasury through a tri-party transaction which is cleared by a clearing bank, and MMMFs are often on the opposite side of the transaction.

who sell (short) the expensive futures contract (step 4) and hedge their position by taking the physical Treasuries from the market, while financing all this activity through repos with their prime brokers (step 5). The very existence of this arbitrage trade demonstrates that the traditional form of intermediation had become more costly for the banks.

In figure 10, the full repo transaction occurs over several steps. In step 5, the prime broker sends cash to the hedge fund, receiving the Treasury as collateral. This lending is bundled as part of the prime brokerage service and is a form of bilateral repo, with the terms of the transaction set by the two parties. The bank may choose to stop here and hold this collateral on its balance sheet, but often this bond will be rehypothecated, this time through the tri-party repo market operating through a clearing bank (step 6). In tri-party repo, the prime broker sends securities and receives cash, while a cash investor (here represented by money market mutual funds) has cash and securities go in the opposite direction (step 8). The clearing bank—the

Figure 11. Leveraged Fund Net Futures Positions: Treasuries



Sources: Commodity Futures Trading Commission (CFTC) and authors' calculations.

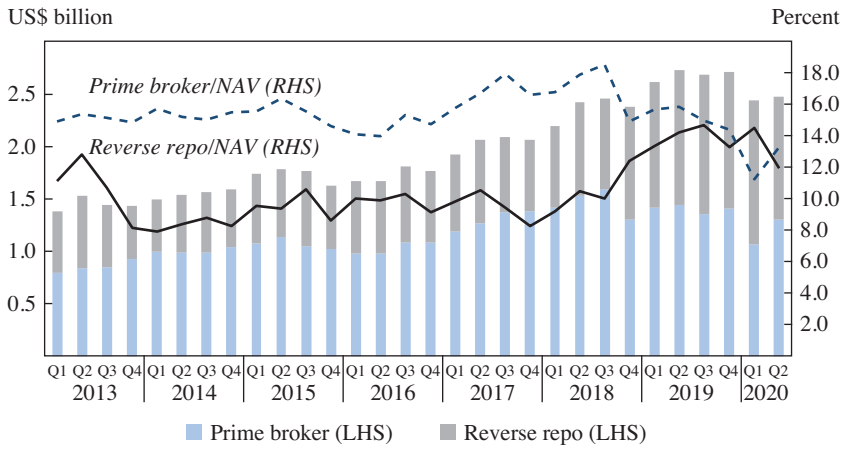
Note: Leveraged funds are financial institutions that use debt to maximize returns on investments. These institutions can invest in Treasuries of different maturities by either shorting the security or going long on the instrument. Shorting involves borrowing securities that one does not own, selling them to another party, and buying them back to return them to the lender at a later date. The goal is to profit off the difference in price if it drops. Investors going long on a security purchase and hold the security with the expectation that it will eventually rise in value. An investor's position with respect to a given security can be net long or net short based on the makeup of its portfolio.

Bank of New York is the only remaining actor in this business—manages the cash and collateral.

While we will refer to the process in figure 10 as "hedge-fund centered intermediation," we note here that intermediation is not the intent of the hedge funds. Instead, these funds are simply doing what they always do—searching out profitable arbitrage opportunities. The profit for these funds is purely from the trade itself, and they have no ability to cross-sell this function with other client-facing activities. Since banks do have such cross-selling opportunities, they have historically been the efficient provider of the intermediation. The fact that banks have removed themselves from part of this chain is a demonstration not of competition directly for their intermediation services but of the substitution of a less direct pathway by agents having a variety of different incentives.

Figure 11 shows some evidence of this switch through the rising participation of hedge funds on the CME Treasury futures market. Even as hedge funds were increasing their participation in futures markets, they were also increasing their repo financing from their prime brokers—the other part of the intermediation. This increase is illustrated in figure 12.

Figure 12. Funding for Qualifying Hedge Funds



Source: SEC Private Fund Statistics.

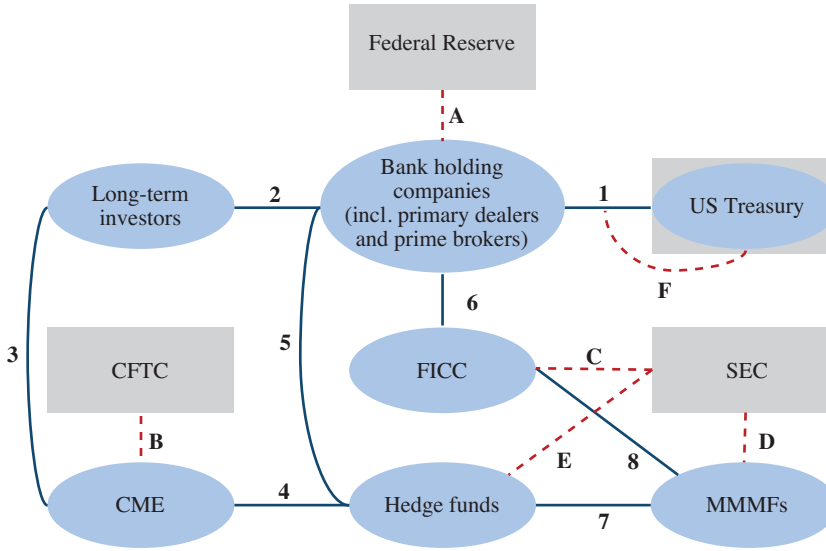
Note: Qualifying hedge funds are those that are required to submit Form-PF with the SEC, and two major sources of funding for such funds are reverse repo or their prime brokers. The net asset value (NAV) is the total value of the fund's assets less its liabilities.

The full intermediation chain includes three different analogues for traditional bank capital requirements: (1) the supplementary leverage ratio imposed on banks for gross repo positions, even with Treasury collateral; (2) the repo haircuts charged by prime brokers in the bilateral market and by the ultimate cash suppliers in the tri-party market; and (3) the initial margins charged by the CME for the futures transactions. This third analogue—initial margin—introduces a new path for incongruity beyond the two seen in the nonprime mortgage market.

In the recent years, regulatory changes and institutional innovations have opened up a new mechanism for the repo transactions. This new mechanism replaces the bilateral/tri-party nexus with a central counterparty, the Fixed Income Clearing Corporation (FICC). This is illustrated in figure 13.

Here, the FICC has taken the place of the tri-party clearing bank in figure 10, which eliminates the need for the initiating bilateral repo between the prime broker and hedge fund. Instead, the prime broker is able to "sponsor" the hedge fund at FICC, and the effective trade between MMMFs and hedge funds occurs at the clearinghouse instead of through the prime broker. In 2017, a rule change by the FICC allowed a broader class of institutions to participate as sponsored members for repo transactions, and

Figure 13. Nonbank Intermediation of Government Securities, with Sponsored Repo



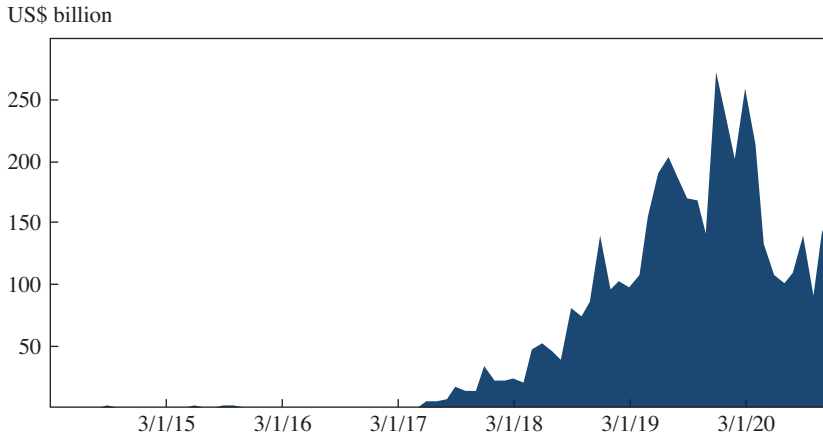
Source: Authors.

Note: The US Treasury sells bonds through auctions to primary dealers—the bank holding companies (1). The Treasury also has regulatory authority of the market (F). In this diagram, hedge funds short the Treasury futures contract and purchase Treasury securities (4). Hedge funds finance this "relative value" trade through repos with their prime brokers (5). In this case, the bilateral/tri-party repo from figure 10 has been replaced by the FICC. The prime broker can sponsor the hedge fund at the FICC, which removes the activity from its balance sheet. MMMFs continue to be on the opposite side of the repo transaction, lending cash for the Treasury securities.

a 2019 rule change expanded eligibility for sponsoring members (Securities and Exchange Commission 2017, 2019). With this broader class of participants, prime broker subsidiaries of BHCs were able to move toward the repo structure shown in figure 13 with many more of their hedge fund clients, and they did so with great speed. By the beginning of the COVID-19 crisis, sponsored repo at FICC grew from negligible pre-2017 levels to nearly $500 billion (Securities and Exchange Commission 2021). This growth did not replace the tri-party/bilateral repo chain shown in figure 10; instead, sponsored repo through FICC has grown dramatically while tri-party repo has remained flat.

For the BHCs, the advantage of sponsored repo is to allow the matching and clearing to occur away from the balance sheet of the bank. Rather than the bilateral/tri-party nexus of figure 10, sponsored repo at FICC had no balance sheet cost to the sponsoring banks, and thus no impact on ESLR requirements. For hedge fund clients, the financing was the same, but there

Figure 14. Money Market Mutual Fund Investments in Repo Cleared by the FICC



Source: Office of Financial Research (OFR).

Note: Sponsored repo is a transaction that allows a dealer to sponsor non-dealer counterparties on the Fixed Income Clearing Corporation's (FICC) cleared repo platform. Repo trades are matched and netted, which has a smaller balance sheet impact for dealers. Changes in 2017 and 2019 expanded eligibility for sponsored and sponsoring members.

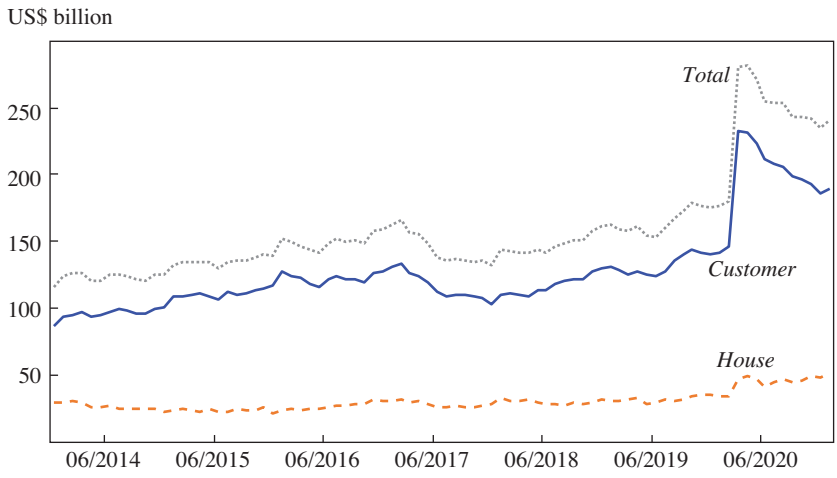
was the additional hassle of dealing with a new counterparty. The fact that the business grew so fast despite this hassle suggests that banks were subsidizing the move.

Once again, it was the MMMFs that provided the majority of funding for this new activity, with provision rising from nothing pre-2017 to over $250 billion in early 2020 (figure 14).

In figures 10 and 13, the Treasury is still selling its securities, which eventually will be delivered to long-term investors upon maturity of the futures contracts. But the middle game has more steps than in figure 9, and the broker-dealer services formerly provided by large banks are now performed by a combination of managed investment pools (hedge funds and MMMFs), facilitated by clearing banks (Bank of New York), futures exchanges (the CME), or central counterparties (FICC). An intermediation system anchored by lightly regulated investment pools can look very different from one anchored by highly regulated banks, and recent years witnessed exactly this shift.

This hedge fund–centered intermediation operates under a very different capital regime than does the simpler bank-centered model. Instead of the 5 percent ESLR requirement from the latter case, the capital securing the hedge fund intermediation is from disparate sources, consisting of

Figure 15. Initial Margin Requirements for Futures



Source: Commodity Futures Trading Commission (CFTC).
Note: For a futures contract, the initial margin is the amount required to be covered with cash or adequate collateral.

(1) margin on the futures short sale, as imposed by the CME; (2) haircuts on the repo transactions, as imposed by the prime broker (for the bilateral step) and by the clearing bank (for the tri-party step); and (3) capital requirements on the parent BHC, to the extent that the bank is unable to net the repo transactions on its balance sheet. In section III, we go into more detail about the mechanics and regulatory requirements of these margin and haircut decisions; for now, it suffices to point out that they are decentralized and uncoordinated, in contrast to the simple ESLR requirement from figure 9. For a BHC considering the least-cost method to provide prime brokerage services, the shifting burdens of these various requirements will heavily influence the location of the activity. And the result of those shifting burdens was to move more of the activity to the indirect hedge fund version of intermediation.

The general malfunction in the Treasury market in March 2020 included several links from the hedge fund intermediation. Of most interest to our project here is the change in initial margins for Treasury futures (figure 15).

While regulators have incorporated more countercyclical features in bank capital requirements since the global financial crisis, initial margin (an analogue for futures) will mechanically be pro-cyclical, as seen here. Despite attempts to push back on such pro-cyclicality, initial margins are model driven, and increases in volatility during a stress event will necessarily work

through the model to increase margins. In the next section, we discuss the challenge of making these margins congruent to capital regulation.

III. Aligning Bank Capital, Repo Haircuts, and Market Margining Requirements

Agreement on the congruence principle would provide the foundation from which to derive policies for managing the contribution of nonbank intermediation to systemic risk. However, as with all policymaking, legal and institutional factors beyond the appeal of the concept will substantially determine how effectively it could be implemented. These factors will be more significant in the United States than in most other key financial jurisdictions because of the famously balkanized organization of financial regulation. At the federal level alone there are three bank regulators and two market regulators. There is no federal regulation of insurance companies, even those with activities ranging far beyond traditional insurance businesses. The Financial Stability Oversight Council (FSOC), created following the global financial crisis, is supposed to coordinate regulatory policy to protect financial stability. But it is structurally flawed and, in any case, possesses little in the way of actual authority.

The challenges can be illustrated using our suggestion of congruent margining, repo haircuts, and capital requirements to ask what would need to be done and, of no small importance, who would do it. Consider the Treasury bond intermediation shown in figure 13 of the previous section. Here, the key regulatory relationships are denoted by letters, with relevant regulation applying to different numbered connections in the diagram. The list below, drawn from that diagram, focuses specifically on capital requirements, where "capital" means not just the rules set by government regulators, but also market-driven haircut and margining practices, which also provide a buffer against losses.

III.A. Federal Reserve Regulation of BHCs

After buying Treasuries at auction (step 1), the banks need to decide how long these Treasuries would be held as inventory on their balance sheets. On a balance sheet, all Treasuries would be subject to a 5 percent ESLR if held at the BHC level.[12] The risk-based capital charge would be zero for bills, but the longer-dated maturities could incur a risk-based capital charge

12. Enhanced Supplementary Leverage Ratio Standards, 79 Fed. Reg. 24528 (May 1, 2014) (codified at 12 C.F.R. pt. 217).

if held on the trading book. In the discussion below, we focus on the 5 percent ESLR charge as a benchmark. It was this charge that—crucially—was waived by the Federal Reserve in the spring of 2020 for one year, thereby freeing up balance sheets for banks to retake this intermediation function (Federal Reserve Board 2020b). Thus, regulatory action was explicitly *countercyclical*; crucially, regulators have discretion to make such countercyclical changes and are not bound by any fixed formula to do so.[13]

III.B. Commodity Futures Trading Commission Regulation of the CME

When hedge funds take over some of the intermediation of Treasuries, they do so by selling Treasury futures in step 4. For such short sales, they must post initial margin beyond the proceeds of the sale and update this margin as prices fluctuate during the life of the contract. For now, we focus on the initial margin, the level of which will be an important input to the total amount of leverage that funds can dedicate to this trade. This initial margin calculation is made by the CME, using models that aim (first approximation) to ensure a 99 percent chance of coverage over a preset horizon (Waldis 2020). The Commodity Futures Trading Commission (CFTC) involvement in model development is high level: it does not review model parameters or set specific levels for key variables. Left to their own devices, standard models will tend to increase initial margin during volatile periods. Notably, the exchange and clearinghouse industry has recognized the danger of pro-cyclicality and has taken steps through a statement of principles to minimize it (Committee on Payment and Settlement Systems and International Organization of Securities Commissions 2012). These steps build a buffer and effectively slow—but do not stop—the pro-cyclical adjustment of initial margin. In March 2020, the CME ultimately increased initial margin right in the midst of the market stress (CME Group 2020). Defenders of this practice can correctly argue that the clearinghouse needs to be protected, since it too is systemically risky. But there is certainly a difference between microprudential protection of a clearinghouse and macroprudential concerns of a liquidity panic. And, under the current rules, there is no regulatory body actively involved in this decision. It is entirely driven by the rules set by the exchange. We return later to a discussion of this important point.

13. The Federal Reserve allowed the waiver to expire in March 2021, apparently without any immediate consequences for bank balance sheets and capital requirements (Federal Reserve Board 2021).

III.C. SEC Regulation of the FICC

The FICC, as a subsidiary of the Depository Trust and Clearing Corporation, is regulated by a self-regulatory organization, the Financial Industry Regulatory Authority (FINRA), which is itself regulated by the SEC. For the sponsored repo shown in steps 6 and 8, FICC sets haircuts and may make future adjustments to collateral, thus acting analogously to the CME for futures. Here, we have even less visibility into the underlying models, and we have no data to tell us whether collateral calls were a source of stress on this part of the market. What we do know is that the haircut and collateral agreements are bespoke with each sponsoring member and thus are market driven. In making those judgments, the FICC and its regulators are naturally concerned about the safety of the clearinghouse itself and have no statutory requirement to consider impacts on overall financial stability. The danger here, as in the CME-CFTC case, is that the initial margins and haircuts in non-stress times will be too low from a macroprudential perspective, incentivizing the intermediation activity to move to these venues. The downside risk will then de facto be absorbed by countercyclical adjustments by the safety net.

III.D. SEC Regulation of MMMFs

The MMMFs are important cash providers in the FICC-sponsored repo market (step 8 of figure 13) and also in the parallel tri-party repo market (step 7 in figure 10). In each of these cases, the MMMFs have handed off their risk to a central counterparty and should be indifferent to specific haircut decisions. But these funds are impacted by regulation in other ways, several of which were relevant for market dysfunction in 2008 and then again in 2020. In particular, MMMF concerns about meeting liquidity rules led them to sharply decrease the maturity of their holdings in March and to pull back on term-repo funding (Eren, Schrimpf, and Sushko 2020).

III.E. SEC Regulation of Hedge Funds

For hedge funds, the long leg of the intermediation comes from the purchase and subsequent repo of physical Treasury securities. While the largest hedge funds are under limited regulatory authority of the SEC (mostly reporting), there are no direct capital requirements for their whole portfolio. Thus, under present regulations, the haircuts paid by hedge funds—either as part of sponsored or bilateral repo—are market driven. This appears to be the most difficult place to achieve congruence, since the

statutory authority to regulate hedge fund capital does not exist and, if it did, would likely just allow that activity to move to another entity.

III.F. US Treasury Authority over the Government Securities Markets

The inclusion of the Treasury Department here may surprise some readers, even those expert in financial stability policy. But the little-known Government Securities Act of 1986 gives the Treasury relevant contingent regulatory authority and, potentially, active leadership of the exercise we contemplate.[14] This legislation was passed in response to the failures of firms that dealt only in government securities and thus escaped regulation by the SEC.[15] It requires the Treasury to adopt rules governing financial responsibility and reporting requirements of government securities brokers and dealers.[16] In practice, the Treasury has exempted from its financial responsibility requirements (covering capital and related measures) any financial institution already subject to the jurisdiction of a market or banking regulator.[17] But its capital regulations—which largely mirror SEC requirements—apply to any freestanding government securities dealer. And it retains the option to adopt additional or more stringent requirements than those imposed by the functional regulators.

How can we get congruence from this morass? To start, we divide the task into three steps. First, the portion of prudential regulatory requirements motivated by concerns other than financial stability aims would be separated out from the effective regulatory charge associated with either holding Treasuries or using them as collateral to obtain funding. Second, the remainder of that regulatory charge would be converted into an initial margin equivalent which, as we suggested, would effectively combine capital and liquidity requirements. Third, that margin equivalent would be applied to the different trading platforms and arrangements, perhaps with adjustments to take account of different risk factors associated with each.

The first two steps would require considerable analysis and, ultimately, regulatory judgment. There will likely be a good bit of debate around what part of the regulatory costs of these transactions for banks is attributable to their special status within the US financial system, either as a whole or individually. Some issues would be fairly clear-cut. Thus, for example,

14. Government Securities Act of 1986, Pub. L. No. 99-571, 100 Stat. 3208 (codified at 15 U.S.C. § 78o–5).

15. United States Code 15 U.S.C. § 77c(a)(2).

16. United States Code 15 U.S.C. § 78o-5(b)(1)(A)–(B).

17. Code of Federal Regulations 17 C.F.R. § 402.1(b)–(c).

the Federal Deposit Insurance Corporation (FDIC) deposit insurance premium that banks must pay on (uninsured) repo liabilities should probably be carved out.[18] Other issues would be considerably less straightforward. In the context of Treasuries, for example, an important question would be whether the appropriate input into the process of converting prudential requirements would be the minimum leverage ratio imposed on all banks or the higher ESLR imposed on the most systemically important banks. The latter is imposed in accordance with the Dodd-Frank principle of more stringent regulation on systemically important banking organizations, based on the greater negative externalities that would follow from their insolvency (and, perhaps, the associated moral hazard).

Similarly, the second step of converting the macroprudential component of the prudential banking regulations into specific minimum margining requirements would be a complicated process involving important, nonobvious decisions. For example, since capital requirements for the largest banks are clearly intended to mitigate microprudential and macroprudential concerns, there is no specific formula to divide the total requirement between these two motivations. However, like the issues raised in the first step, it seems no more complicated or judgmental than many existing financial regulatory efforts, including development of the very regulatory capital standards that would be the starting point for this exercise. Moreover, to a considerable extent the work here would be heavily front-loaded. Once regulators made some of the key threshold judgments, revisions to the original regulation or application of the congruence principle to other shadow banking activities should be somewhat more straightforward.

There would need to be agreement on methodologies for the first and second steps among all financial regulatory agencies whose legal authorities would be needed to implement congruent margins across markets and platforms. Here is where the balkanization of US financial regulatory authority becomes a consideration (Yadav 2019). Under current law and practice, achieving congruence for transactions collateralized with Treasuries would involve six agencies: the US Treasury itself; the three federal banking agencies—the Federal Reserve, the Office of the Comptroller

18. In fact, there is a good argument for requiring all forms of short-term financing to bear a fee reflecting the implicit systemic risk insurance provided by the Federal Reserve. However, the extension of the FDIC premium requirement to the uninsured liabilities of BHCs was essentially a way to augment the insurance fund without raising premia on the insured deposits of depository institutions, not a fee calibrated to insurance provided by the Federal Reserve.

of the Currency (OCC), and the FDIC; the CFTC; and the SEC. While the SEC and CFTC obviously do not set capital requirements for banks, congruence will be achieved only if all agencies agree on the amount of resiliency that will align margining requirements with bank regulatory requirements. And, although regulatory relationship A in figure 13 refers to BHCs and thereby originates only at the Federal Reserve, national banks within BHCs are directly regulated by the OCC, and state-chartered banks that are not members of the Federal Reserve System are directly regulated by the FDIC. Under long-standing practice the three federal banking agencies jointly adopt capital requirements applicable to BHCs and all insured depository institutions.[19]

Needless to say, this would not be a nimble process, as evidenced by experience with mandatory joint rule making under the Dodd-Frank Act. These efforts have generally been slow moving, especially in rule making involving both banking and market regulators. The sheer complexity of coordinating agreement among six sets of agency staffs and principals (a total of as many as twenty-three Senate-confirmed presidential appointees) can make for cumbersome regulations that reflect sometimes awkward compromises. Of equal or, at times, greater importance has been the difference in perspectives and missions of regulators. The original missions of the market regulatory agencies, which gave rise to their institutional cultures, were those of investor protection and operationally well-functioning markets. While financial stability has always been at least a background concern of bank regulators, it was considerably more peripheral to the market regulators. During the protracted interagency negotiations over Dodd-Frank mandated rules, including the very relevant rule on minimum swap margins, these differences in perspective—especially at the SEC—could be substantial hurdles to agreement.

As shown by the swap margin rule making, for example, it can be done. But the prospect of the prolonged and difficult process that eventually produced that and other Dodd-Frank rules could present an additional hurdle to realizing a congruent margining regime.[20] Unlike the joint regulations

19. The joint rule making on capital and certain other prudential standards is a voluntary undertaking of the banking agencies, in contrast to the many instances of joint rule making required by the Dodd-Frank Act, as discussed in the next footnote.

20. Some of the joint rule making required by Dodd-Frank has not been completed more than a decade after that law was enacted. Notably, there is still no rule on limiting bank incentive compensation arrangements that encourage inappropriate risk-taking, as required by section 956 of Dodd-Frank. While the problems in developing a rule on this immensely complicated topic are not wholly attributable to differences in policy perspective and agency practices on compliance, they have certainly played a role.

on swaps and the Volcker Rule, which were mandated by Congress, here there could be an initial hurdle in gaining agreement among the agencies to undertake the effort in the first place. Still, we think there are grounds for optimism, albeit guarded optimism. First, the set of agency principals appointed by President Biden, including those at the market regulatory agencies, may well share a stronger inclination toward acting on NBFI activities than their predecessors. Second, the Treasury's authority under the Government Securities Act significantly increases its leverage, both in initiating and driving the process.

Unlike the many macroprudential regulatory issues for which the Treasury lacks authority beyond its hortatory role as head of the FSOC, here the Treasury has a key, possibly central, role. The Government Securities Act requires it to adopt rules that "provide safeguards with respect to the financial responsibility and related practices" of government securities brokers and dealers.[21] To date, the Treasury has generally taken a minimalist approach to its regulatory role, adopting capital and other requirements for dealers trading exclusively in government securities that parallel those adopted by the SEC for dealers more generally. But in the Government Securities Act Congress stated its purposes in far-reaching terms, to "provide for the integrity, stability, and efficiency of . . . transactions" in government securities and "to protect investors and to insure the maintenance of fair, honest, and liquid markets in such securities."[22] While the direct motivation for the act was the failure of certain firms that had dealt only in government securities and were thus exempt from SEC regulation, the law gave the Treasury both a broad mandate to protect the government securities market and the regulatory tools to adapt to evolving risks to the integrity of that market.[23] Thus, while a Treasury regulation requiring minimum haircuts would be a new use of the authority, it would be entirely consistent with the text and purposes of the law. In this context, it would be sensible for the Treasury to dust off its authority under the Government Securities Act and to exercise that authority with an eye to problems in Treasury markets that did not exist when the act was passed.

The prospect of Treasury rules that would augment or displace the rules of functional regulators pertaining to government securities transactions should provide considerable incentive for the other agencies to cooperate

21. United States Code 15 U.S.C. § 78o–5(b)(1)(A).

22. Government Securities Act of 1986, Pub. L. No. 99-571, § 1, 100 Stat. 3208, 3208 (codified at 15 U.S.C. § 78o–5).

23. For a discussion of current problems in the Treasury market, see Liang and Parkinson (2020).

in the process we sketch out.[24] Otherwise, the Treasury might dictate capital or margin rules for dealing in Treasuries that sit uncomfortably with existing regulatory regimes. On the other hand, since the Government Securities Act places most responsibility for implementation and enforcement of its regulations in the hands of the financial regulators, the Treasury needs their cooperation and expertise to attain its desired regulatory outcome. So, we hope, all six agencies have reason to support this collective process. Indeed, a good prelude to the process would be formal endorsement by the FSOC agencies of the congruence principle and a public commitment to cooperative solutions to the risks created by NBFI activities.

One final point on this subject—while the Treasury's largely unexercised authority gives it the ability to initiate and drive a collaborative effort to achieve congruency in Treasury-backed repo markets, it also means that the Treasury could impede any such process. For example, Treasury officials could be skeptical of congruence efforts that might incrementally decrease the attractiveness of holding Treasuries, thus increasing the government's debt servicing costs. In such cases, it would be awkward at best for the other agencies to use their broad authority over banks and general broker-dealers to impose congruent margins and haircuts.

The third step is certainly not more analytically difficult than the first two steps. But it introduces issues of legal authority, organizational capacity, and agency traditions that both complicate the execution of the first two steps and raise potential trade-offs between theoretically preferable regulatory features and practical questions of administrability.

After agreement on the methodologies in the first two steps, regulators would need to adopt regulations binding both central counterparties and significant non-centrally cleared Treasury repo activity by non-prudentially regulated actors. As for the central counterparties, the CFTC would establish minimum margins for Treasuries futures trading on the CME. Either directly, or through FINRA, the SEC would establish minimum haircuts for FICC (regulatory relationships B and C, respectively, in figure 13). A significant institutional choice will need to be made during the step 3 process of applying the initial margin requirements in different market contexts (i.e., in relationships B, C, E, and possibly D): Would regulators themselves calculate the minimum initial margin or haircut for CME, FICC, and other central counterparties? Current practice, as included in formally adopted, binding regulations, is that the market regulators adopt principles-based regulations for central counterparties, which themselves determine

24. United States Code 15 U.S.C. § 78o–5(b)(1)(A)–(B).

specific margin requirements. Given the complexities of central clearing and the risk-related differences that may exist among different clearing entities, there is reason for this practice. Yet the experience of March 2020 raises the issue of whether regulatory oversight of central counterparties' margining practices is, or can be, sufficiently rigorous. If effective monitoring is not possible, then specific mandatory margins or margining formulas might be necessary.

While trade-offs between conceptually desirable regulatory tailoring and administrability are regularly made in financial regulation, the difference in perspective between market regulators and banking regulators could cause additional coordination difficulties. The latter, chastened by the failures of bank modeling and risk management in the years leading up to the global financial crisis, may be more skeptical of frameworks that contain a good bit of what is effectively self-regulation.

Under current statutory authority, the other task for regulators—covering significant Treasury repo activity by nonbanks that is not centrally cleared—would need to be addressed somewhat differently. In addition to current activity falling into this category, one could expect that in the absence of comprehensive regulation certain market actors would see opportunities for arbitrage. Actors falling outside applicable regulatory perimeters could be incentivized to amass cash pools that could engage in Treasury-backed repo lending without the minimum haircut or capital requirements applicable in transactions involving central clearing or banks.

No agency, including the Treasury, has authority to impose minimum margining requirements on such activity directly—that is, on a transaction basis. With respect to entities that are already registered as dealers, whether or not part of a BHC, the SEC has authority to require them to apply minimum margins when they extend credit against Treasuries. However, the Treasury has authority over all government securities dealers, defined as "any person engaged in the business of buying and selling government securities for his own account."[25] The question would then be whether any other entity regularly engaged in substantial amounts of repo transactions is, within the terms of the securities laws, "engaged in the business" of buying and selling government securities.

The statutory definition of "government securities dealer" appears broad enough to capture any existing or new entity that participated in substantial Treasury-backed repo activity. The definition has not been elaborated upon by the courts. However, the definition of "dealer" subject to SEC

25. Securities Exchange Act of 1934 § 3(a)(44), 15 U.S.C. § 78c(a)(44).

jurisdiction—on which the government securities dealer definition was modeled—has been broadly construed. Courts have found the statutory definition of dealer—"any person engaged in the business of buying and selling securities"—to cover a range of situations in which market actors regularly involved themselves in purchases and sales.[26]

The Treasury would need to establish minimum haircuts to cover non-centrally cleared transactions by dealers not already subject to the jurisdiction of a financial regulatory agency. Like the Treasury's capital requirements for such entities, its haircut requirements would presumably parallel those developed by the SEC. Because the SEC has broad authority to set conditions for the operation of MMMFs, it could either impose haircuts directly through regulatory relationship D (figure 13) or leave those funds subject to requirements that the Treasury establishes directed at the broader market in regulatory relationship F.

While, as a technical matter, the banking agencies would not need to adopt their own regulations as part of the initial effort to conform margining and haircut practices to the resiliency standards implicit in capital regulation, there are two ways in which they, too, might modify their regulations in implementing the congruence principle. First, of course, changing market practices and regulatory assessments of risk could counsel changes in capital, as well as margining, requirements. Second, since minimum margins would by definition apply to the trading of prudentially regulated entities on central clearing platforms, the banking agencies might want to adjust some of their prudential regulations to take account of the stricter margining requirements. Similarly, since banks are exempted from the universe of dealers subject to SEC and Treasury regulation, the banking agencies may want to consider imposing the same margin requirements on repo lending by banks and adjusting capital requirements to take account of this change.

The mREIT case that served as our first example could be addressed through a comparable process, though here the role of catalyst would be played by the Federal Reserve. The Securities Exchange Act of 1934 authorizes the Fed to set minimum margins on most securities financing transactions backed by collateral other than government securities, which were

26. See *SEC v. River N. Equity LLC*, 415 F. Supp. 3d 853, 858–59 (N.D. Ill. 2019) (noting that the Exchange Act "broadly defines 'dealer'" because the act's dealer registration requirement is "of the utmost importance in effecting the purposes of the Act"); see also *SEC v. Fierro*, Civil Action No. 20-2104 (MAS) (DEA), 2020 WL 7481773, at *3–*4 (D.N.J. Dec. 18, 2020) and *SEC v. Keener*, 1:20-CV-21254, 2020 WL 4736205, at *3–*5 (S.D. Fla. Aug. 14, 2020).

excluded from the act's coverage.[27] In practice, the Fed has delegated that authority to the market regulators, just as the Treasury has for government securities. But, as with the Treasury for government securities, the Fed has leverage over the market regulators and residual authority that can be used to drive a congruence exercise. Indeed, in 2015 the Fed anticipated use of that authority in supporting an international agreement to require minimum haircuts on non-centrally cleared securities financing transactions using any collateral other than government securities (Financial Stability Board 2015). That agreement, whose implementation has been repeatedly delayed by the Fed and the other members of the Financial Stability Board, could pave the way for efforts to extend congruence efforts internationally.

In fact, this process would be somewhat simpler than the Treasuries exercise, since neither the Treasury nor the CFTC regulates any of the principal actors. Thus the first two analytic steps would involve fewer agencies. As illustrated in figure 4, the key implementation action in the third step would be on line F, where the SEC would require minimum margining practices by the broker-dealers that are the repo counterparties of the mREITs. As with the Treasury example, the Federal Reserve would have the option of adopting its own regulation or allowing the SEC—acting in the shadow of the Fed's authority—to make appropriate modification of its regulatory requirements for dealers. Again, were market actors to form a cash pool and undertake repo activity with the mREITs, the securities law definition of "dealer" is likely broad enough to capture that intermediary.

IV. Conclusion

The prospect of protracted interagency negotiations is hardly encouraging for those, including us, who believe that containment of systemic risk outside the prudentially regulated sector is both important and long overdue. But what are the alternatives? We can think of two.

One alternative is to eschew formal efforts at coordination and to rely on each agency to address risks arising within its usual regulatory domain. This describes the status quo, though it is reasonable to assume that the COVID-19 market turmoil and the arrival of Biden-appointed leadership will push each agency toward more rigor. Still, the likelihood of divergence

27. The authority is given in section 7 of the Securities Exchange Act of 1934, 15 U.S.C. § 78g. The Fed's implementation of that authority is contained within its Regulation T, 17 C.F.R. pt. 220. There are some statutory limits on this authority, which probably would not be too consequential for our purposes.

in the effective regulatory charges on similarly risky activities when conducted by different kinds of intermediaries will reproduce opportunities for regulatory arbitrage. That, of course, is much of what has motivated our congruence principle in the first place.

The second alternative would be to expand agency regulatory authority, concentrate it in a smaller number of regulatory actors, or both. One approach would be a version of proposals that are periodically advanced to reduce the number of financial regulatory agencies. Whatever the merits of consolidating and enhancing agency authority to counteract systemic risk, near-term prospects for such legislation are at best modest. In the aftermath of the global financial crisis, of which financial intermediaries were a principal cause, the only institutional changes were the elimination of the much-criticized Office of Thrift Supervision, the creation of the FSOC with its cumbersome organization and limited authority, and the creation of a new agency with the sole mission of consumer financial protection. Indeed, the approach in Dodd-Frank—allocating additional regulatory jurisdiction among agencies and then requiring joint rule making—increased, rather than diminished, coordination needs.

Prospects for a major reorganization of the relevant governmental agencies directed at more effective control of systemic risk are likely even lower following the COVID-19 crisis than they were after the global financial crisis. In the spring of 2020 the vulnerabilities of nonbank intermediaries amplified stress but did not create it. Other legislative priorities, the interests of some groups in maintaining a balkanized regulatory regime, and broader policy concerns about further concentrations of regulatory authority combine to make such legislation a long shot for the foreseeable future.[28] Finally, it is worth noting that more concentrated authority would not eliminate the need to tailor margining and other regulatory requirements to the varying risk characteristics of trading markets in which similar financial

28. Here are two examples observed by one of the authors who was involved in the process that led to the Dodd-Frank Act: First, during debate over the Dodd-Frank Act, proposals to consolidate bank regulatory authority in two agencies (one for national banks and one federal regulator for state banks) were opposed by many medium-sized and smaller banks, which wanted to retain the option of switching primary federal regulators. Second, key members of the two congressional agriculture committees have in the past strongly resisted transferring to the SEC the CFTC's authority over financial futures (as opposed to futures for physical commodities—the original motivation for the CFTC). While commentators (and legislators on the banking committees) have advanced policy arguments for doing so, giving up such an influential oversight and legislative role is not a natural instinct of most legislators.

activity is being conducted. That is, even under a single regulatory authority, the same factors that argue for congruence rather than equivalence will require the expertise currently located in the SEC and CFTC.

There are, of course, many other worthwhile suggestions that have been offered in response to the growth of NBFI contributions to systemic risk. Unfortunately, many of these suggestions would likely, or certainly, require new legislation. So, for example, existing legal authorities may not extend to requiring all significant financial intermediaries to participate in a system-wide stress test. A proposal that any intermediary receiving liquidity support during stress periods thereafter become subject to some form of prudential regulation has some intuitive appeal, but it would surely require congressional action. Here again, while we can remain hopeful, we are not sanguine about near-term chances for legislation to extend the prudential regulatory perimeter.

In sum, we are very much aware of the institutional and practical constraints on realizing in practice the congruence that we find so compelling in principle. We would be delighted to learn of other approaches that would be more efficient and effective. But, in the absence of such ideas, and with the reality of substantial hurdles to legislative solutions, we believe that interagency processes in which the Treasury and Federal Reserve have the legal authority to take leadership is superior to the currently available alternatives.

ACKNOWLEDGMENTS We thank Viral Acharya, Jan Eberly, Gary Gorton, Howell Jackson, Hyun Song Shin, Jeremy Stein, and Joshua Younger for helpful comments and conversations. We are also grateful to staff at the financial regulatory agencies for helping us understand better the nuances of current regulatory practice. Alison Brockman, Mallory Dreyer, and Sharon Nunn provided excellent research assistance.

References

Acharya, Viral V., and T. Sabri Öncü. 2012. "A Proposal for the Resolution of Systemically Important Assets and Liabilities: The Case of the Repo Market." Paper prepared for Central Banking: Before, During and After the Crisis, Federal Reserve Board, Washington, March 23–24, 2012. http://pages.stern.nyu.edu/~sternfin/vacharya/public_html/Resolution%20of%20Systemically%20Important%20Assets%20and%20Liabilities_Sabri_VA_10.pdf.

Acharya, Viral V., Philipp Schnabl, and Gustavo Suarez. 2013. "Securitization without Risk Transfer." *Journal of Financial Economics* 107, no. 3: 515–36.

Bank of England. 2020. *Financial Stability Report*. London: Financial Policy Committee. https://www.bankofengland.co.uk/-/media/boe/files/financial-stability-report/2020/august-2020.pdf.

Barth, Daniel, and Jay Kahn. 2020. "Basis Trades and Treasury Market Illiquidity." Brief 20–01. Washington: US Department of the Treasury, Office of Financial Research. https://www.financialresearch.gov/briefs/files/OFRBr_2020_01_Basis-Trades.pdf.

Barth, Daniel, and Jay Kahn. 2021. "Hedge Funds and the Treasury Cash-Futures Disconnect." Working Paper 21–01. Washington: US Department of the Treasury, Office of Financial Research. https://www.financialresearch.gov/working-papers/files/OFRwp-21-01-hedge-funds-and-the-treasury-cash-futures-disconnect.pdf.

CME Group. 2020. "CME Clearing's Margining Practices: CoVid Pandemic." Presented at the Global Markets Advisory Committee Meeting, Teleconference, December 17. https://www.cftc.gov/media/5496/GMAC_121720CME/download.

Committee on Payment and Settlement Systems and International Organization of Securities Commissions. 2012. "Principles for Financial Market Infrastructures." Basel: Bank for International Settlements. https://www.bis.org/cpmi/publ/d101a.pdf.

COVID-19 Market Impact Working Group. 2020. *Experiences of US Money Market Funds during the COVID-19 Crisis*. Washington: Investment Company Institute. https://www.sec.gov/comments/credit-market-interconnectedness/cll10-8026117-225527.pdf.

Duffie, Darrell. 2018. "Financial Regulatory Reform after the Crisis: An Assessment." *Management Science* 64, no. 10: 4835–57.

Duffie, Darrell. 2020. "Still the World's Safe Haven? Redesigning the U.S. Treasury Market after the COVID-19 Crisis." Working Paper 62. Washington: Brookings Institution, Hutchins Center on Fiscal and Monetary Policy. https://www.brookings.edu/wp-content/uploads/2020/05/WP62_Duffie_v2.pdf.

Eren, Egemen, Andreas Schrimpf, and Vladyslav Sushko. 2020. "US Dollar Funding Markets during the Covid-19 Crisis—the International Dimension." BIS Bulletin 15. Basel: Bank for International Settlements. https://www.bis.org/publ/bisbull15.pdf.

Federal Reserve Board of Governors. 2020a. "Federal Reserve Announces Extensive New Measures to Support the Economy." Press Release. Washington. https://www.federalreserve.gov/newsevents/pressreleases/monetary20200323b.htm.

Federal Reserve Board of Governors. 2020b. "Federal Reserve Board Announces Temporary Change to Its Supplementary Leverage Ratio Rule to Ease Strains in the Treasury Market Resulting from the Coronavirus and Increase Banking Organizations' Ability to Provide Credit to Households and Businesses." Press Release. Washington. https://www.federalreserve.gov/newsevents/pressreleases/bcreg20200401a.htm.

Federal Reserve Board of Governors. 2020c. *Financial Stability Report*. Washington. https://www.federalreserve.gov/publications/files/financial-stability-report-20201109.pdf.

Federal Reserve Board of Governors. 2021. "Federal Reserve Board Announces That the Temporary Change to Its Supplementary Leverage Ratio (SLR) for Bank Holding Companies Will Expire as Scheduled on March 31." Press Release. Washington. https://www.federalreserve.gov/newsevents/pressreleases/bcreg20210319a.htm.

Financial Stability Board. 2015. "Transforming Shadow Banking into Resilient Market-Based Finance: An Overview of Progress." Progress Report. Basel. https://www.fsb.org/2015/11/transforming-shadow-banking-into-resilient-market-based-finance-an-overview-of-progress/.

Financial Stability Board. 2020a. "Holistic Review of the March Market Turmoil." Basel. https://www.fsb.org/wp-content/uploads/P171120-2.pdf.

Financial Stability Board. 2020b. "Global Monitoring Report on Non-Bank Financial Intermediation 2020." Basel. https://www.fsb.org/wp-content/uploads/P161220.pdf.

Financial Stability Oversight Council. 2020. *Annual Report*. Washington. https://home.treasury.gov/system/files/261/FSOC2020AnnualReport.pdf.

Fleming, Michael, and Francisco Ruela. 2020. "Treasury Market Liquidity during the COVID-19 Crisis." Blog post, April 17, Liberty Street Economics, Federal Reserve Bank of New York. https://libertystreeteconomics.newyorkfed.org/2020/04/treasury-market-liquidity-during-the-covid-19-crisis.html.

Gete, Pedro, and Michael Reher. 2021. "Mortgage Securitization and Shadow Bank Lending." *Review of Financial Studies* 34, no. 5: 2236–74.

Gorton, Gary, and Andrew Metrick. 2010. "Regulating the Shadow Banking System." *Brookings Papers on Economic Activity*, Fall, 261–312.

Gorton, Gary, and Andrew Metrick. 2012. "Securitized Banking and the Run on Repo." *Journal of Financial Economics* 104, no. 3: 425–51.

Government Accountability Office. 2016. "Mortgage-Related Assets: Capital Requirements Vary Depending on Type of Asset." Report to the Chairman, Committee on Banking, Housing, and Urban Affairs, US Senate GAO-17-93. Washington. https://www.gao.gov/assets/gao-17-93.pdf.

Haddad, Valentin, Alan Moreira, and Tyler Muir. 2020. "When Selling Becomes Viral: Disruptions in Debt Markets in the COVID-19 Crisis and the Fed's

Response." Working Paper 27168. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27168.

Hoffman, Liz, and Gregory Zuckerman. 2020. "RBC Seeks Fire-Sale Buyers for Seized Mortgage Debt." *Wall Street Journal*, March 25. https://www.wsj.com/articles/rbc-seeks-fire-sale-buyers-for-seized-mortgage-debt-11585159333.

International Monetary Fund. 2018. *Global Financial Stability Report: A Decade after the Global Financial Crisis: Are We Safer?* Washington: Author. https://www.elibrary.imf.org/doc/IMF082/25319-9781484375594/25319-9781484375594/Other_formats/Source_PDF/25319-9781484376829.pdf.

International Monetary Fund. 2020. *Global Financial Stability Report: Bridge to Recovery.* Washington: Author. https://www.imf.org/en/Publications/GFSR/Issues/2020/10/13/global-financial-stability-report-october-2020.

Jackson, Howell E. 1999. "Regulation in a Multisectored Financial Services Industry: An Exploratory Essay." *Washington University Law Quarterly* 77, no. 2: 319–97.

Kim, You Suk, Steven M. Laufer, Karen Pence, Richard Stanton, and Nancy Wallace. 2018. "Liquidity Crises in the Mortgage Market." *Brookings Papers on Economic Activity*, Spring, 347–428.

Klingler, Sven, and Suresh M. Sundaresan. 2020. "Diminishing Treasury Convenience Premiums: Effects of Dealers' Excess Demand in Auctions." Working Paper 3556502. Social Science Research Network, March 18. https://doi.org/10.2139/ssrn.3556502.

Kothari, S. P., Dalia Blass, Alan Cohen, and Sumit Rajpal. 2020. *U.S. Credit Markets Interconnectedness and the Effects of the COVID-19 Economic Shock*. Washington: Securities and Exchange Commission, Division of Economic and Risk Analysis. https://www.sec.gov/files/US-Credit-Markets_COVID-19_Report.pdf.

Li, Lei, Yi Li, Marco Macchiavelli, and Xing (Alex) Zhou. 2020. "Runs and Interventions in the Time of COVID-19: Evidence from Money Funds." Working Paper. Social Science Research Network, December 29. https://doi.org/10.2139/ssrn.3607593.

Liang, Nellie, and Pat Parkinson. 2020. "Enhancing Liquidity of the U.S. Treasury Market Under Stress." Working Paper 72. Washington: Brookings Institution, Hutchins Center on Fiscal and Monetary Policy. https://www.brookings.edu/wp-content/uploads/2020/12/WP72_Liang-Parkinson.pdf.

Maloney, Tom. 2020. "After $50 Billion of Losses, No One Comes to Save Mortgage Market with $50 Billion of Losses." Bloomberg, April 3.

Morris, Stephen, and Hyun Song Shin. 2008. "Financial Regulation in a System Context." *Brookings Papers on Economic Activity*, Fall, 229–74.

Pellerin, Sabrina R., Steven J. Sabol, and John R. Walter. 2013. "MBS Real Estate Investment Trusts: A Primer." *Economic Quarterly* 99, no. 3: 193–227.

Quarles, Randal. 2018. "Liquidity Regulation and the Size of the Fed's Balance Sheet." Speech given at Currencies, Capital, and Central Bank Balances: A Policy Conference, Stanford University, Stanford, Calif., May 4.

Scaggs, Alexandra. 2020. "These Two Mortgage REITs Are Falling after a Spectacular Rally." *Barron's*, June 9.

Schrimpf, Andreas, Hyun Song Shin, and Vladyslav Sushko. 2020. "Leverage and Margin Spirals in Fixed Income Markets during the Covid-19 Crisis." *BIS Bulletin* no. 2. Basel: Bank for International Settlements. https://www.bis.org/publ/bisbull02.pdf.

Securities and Exchange Commission (SEC). 2017. "Order Approving a Proposed Rule Change, as Modified by Amendment No. 1, to Expand the Types of Entities That Are Eligible to Participate in Fixed Income Clearing Corporation as Sponsored Members and Make Other Changes." Washington: Author. https://www.sec.gov/rules/sro/ficc/2017/34-80563.pdf.

Securities and Exchange Commission (SEC). 2019. "Order Approving a Proposed Rule Change to Expand Sponsoring Member Eligibility in the Government Securities Division Rulebook and Make Other Changes." Washington: Author. https://www.sec.gov/rules/sro/ficc/2019/34-85470.pdf.

Securities and Exchange Commission (SEC). 2021. "Self-Regulatory Organizations; Fixed Income Clearing Corporation; Notice of Filing of Proposed Rule Change to Add the Sponsored GC Service and Make Other Changes." Washington: Author. https://www.sec.gov/rules/sro/ficc/2021/34-92014.pdf.

Stein, Jeremy C. 2013. "Overheating in Credit Markets: Origins, Measurement, and Policy Responses." Speech given at Restoring Household Financial Stability after the Great Recession: Why Household Balance Sheets Matter, Federal Reserve Bank of St. Louis, St. Louis, Mo., February 7. https://www.federalreserve.gov/newsevents/speech/files/stein20130207a.pdf.

Waldis, Matthew. 2020. "Understanding Margin Changes." Chicago: CME Group. https://www.cmegroup.com/content/cmegroup/en/education/articles-and-reports/understanding-margin-changes.html.

Yadav, Yesha. 2019. "The Failed Regulation of U.S. Treasury Markets." Vanderbilt Law Research Paper 19-12. Social Science Research Network, April 19. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3365829.

Comments and Discussion

COMMENT BY

VIRAL ACHARYA Metrick and Tarullo introduce the notion of a "congruence principle" in regulating financial risk-taking in order to limit the unchecked buildup of risks in nonbank financial intermediaries when risks are regulated in parts of the financial sector, notably at commercial banks and bank holding companies. The idea behind congruence (not necessarily identity) in financial regulation for different parts of the financial sector is that the externality margin by which risk-taking contributes to systemic risk—a collapse of the entire financial sector or a significant portion of it—should be equalized across different parts; in other words, the private gains from undertaking risks should be offset by regulatory costs that reflect the systemic risk contribution of undertaking such risks. Where such offsets are missing, systemic risk would simply migrate to the lightly regulated parts of the financial sector.

Metrick and Tarullo give examples of the growth of non-prime mortgage finance in the United States prior to the global financial crisis and the postcrisis reshuffling of holdings of the US Treasury securities as leading case studies of how risks move from regulated to unregulated parts once regulations are designed, with specific suggestions in these settings of how to apply the congruence principle.

I find Metrick and Tarullo's contribution a timely one, as a decade plus after the global financial crisis risks have begun to proliferate in nonbank financial intermediaries (and even in corporate balance sheets), as witnessed in March 2020 at the onset of the COVID-19 pandemic. It is a reminder that if the approach to financial regulation is not system-based, then we risk facing similar risk imbalances as had developed in the buildup to the global financial crisis, with inevitable financial fragility down the line.

My observations and suggestions on their important contribution are as follows.

CONGRUENCE PRINCIPLE Is the congruence principle that Metrick and Tarullo advocate different from the well-accepted notion that a sound approach to financial sector risk-taking should "regulate by function rather than form"? As my colleague at NYU Stern, Larry White, puts it beautifully in his "waterfall theorem of regulatory arbitrage": *risk travels within the financial sector until it reaches the balance sheet that has the lowest regulatory capital requirement to hold it!* Perhaps when the congruence principle is applied to financial institutions, the two approaches seem the same; however, there may be value to the notion of the congruence principle when applied at the level of financial instruments. Let me elaborate.

It seems to me that one rationale for using the term "congruence principle" would be if the authors were to expand the scope of its definition. The long history of regulatory arbitrage and financial fragility suggests that regulation needs to be harmonized not just across institutional forms but also across financial products (assets, liabilities, etc.) when the latter are similar. Consider the authors' own example of holding mortgages on bank portfolios versus repackaging them as AAA-rated mortgage-backed securities because of the substantial arbitrage in capital requirements due to differences in regulatory risk weights on mortgages versus mortgage-backed securities. Often, such discrepancies combine with regulatory arbitrage via institutional forms to create a complex web of financial transactions that serve no purpose other than being a runaround of regulatory requirements.

Several other examples come to mind, notably capital charges for loans versus those guaranteed by an AAA-rated counterparty, epitomized in the risk-taking by AIG, Inc., that led to its eventual collapse, and capital treatment of liquidity guarantees based on the maturity being below one year versus more than one year, assuming less than one year is necessarily for working capital requirements, when this interpretation was abused for providing guarantees to asset-backed commercial paper held in shadow banking (Acharya, Schnabl, and Suarez 2013).

My suggestion, therefore, would be to adopt this broader definition for the congruence principle as applicable to financial instruments and not just across financial institutions; simply regulating by function rather than form could lead to excess pressure to compromise regulation by violating the congruence principle across systemically important assets and liabilities by repackaging risks through financial engineering.[1]

1. See Acharya and Öncü (2013) for a definition of systemically important assets and liabilities, with a specific application to sell and repurchase (repo) contracts.

BANKING AND SHADOW BANKING LINKAGES When it comes to shadow banking and banking, it is almost always the case that shadow always touches the feet. While the banking and shadow banking linkage does come through where Metrick and Tarullo discuss how regulatory arbitrage simply transfers problems around the financial sector when eventual shocks hit, it would be better to recognize explicitly that banking is more often than not connected to shadow banking. The connections take several forms: explicit guarantees (lines of credit to support commercial paper); implicit guarantees (commercial bank support of specialized investment vehicles or structured investment vehicles prior to the global financial crisis); flow of funds (freezing corporate bond markets can trigger corporate drawdowns on bank credit lines, as occurred in March 2020); and information contagion and interconnectedness.

Such discussion would help bolster the case substantially for getting into the cracks of the financial sector with the congruence principle, as leaving them open in fact threatens commercial banking, which remains the core of payments and settlement systems, deposit provision, and so on.

Equally important, I wonder if it is time to recognize that central banks have essentially embraced head on the idea of being the buyer/market maker of last resort to systemically important markets beyond just being the traditional lender of last resort to banks. This approach was initiated during the global financial crisis and was most recently deployed across the board, including for risky corporate bonds, in the aftermath of the COVID-19 pandemic. This recognition necessitates a system-wide approach that respects the congruence principle across institutions and instruments; in other words, it has now become a no-brainer that we can no longer support any case for not moving *all* large or important markets to centralized platforms for trading and clearing with the necessary transition costs. Central banks have expanded the safety net substantially and likely irreversibly, so the focus must be on ensuring private insurance in all contracts that are the beneficiaries. Metrick and Tarullo's congruence principle can be made the benchmark for ensuring such private insurance is required in a comprehensive manner across different parts of the financial sector.

POLITICAL ECONOMY OF DOMESTIC AND INTERNATIONAL REGULATORY STANDARDS Another thesis states that it is not so much that regulators aren't aware of the congruence principle or of regulating by function rather than form but that the political economy of regulation induces specific regulators to either guard or give way on regulation of their turf while compromising or turning a blind eye on what is outside their turf. Under political pressures and compulsions, regulators may value short-term growth over

long-run financial stability or let the excess occur immediately outside of their sphere of influence. It is thus important to think through arrangements for regulatory decision making that make it robust to political economy pressures. How can international regulatory arbitrage be prevented, as this is what has ultimately caused some banking regulations to weaken globally? Can the Bank for International Settlements or the Financial Stability Board be charged to adopt the congruence principle across institutions and instruments? Would the adoption of the congruence principle simply transform the political economy problem to a complete race to the bottom in terms of the *level* of financial sector regulation (for instance, reduce the level of capital requirements)? These issues appear worthy of discussion while thinking through implementation of the congruence principle in practice.

One other recommendation is that Metrick and Tarullo can lay out clearly the difference between congruence and identity. Is it possible to state some principle for understanding this difference? If not, I am concerned this difference may evolve into a case by case exception, which is precisely how regulatory arbitrage is enforced by financial institutions and their lobbyists in the first place; in particular, they rely on regulatory discretion and its vulnerability to demands based on exceptionalism for relaxation of rules in specific segments of the financial sector.

WAYS TO IMPLEMENT THE CONGRUENCE PRINCIPLE To this end, here are some practical approaches to implementation that Metrick and Tarullo might consider in their future drafts and in their efforts to improve financial sector regulation.

Can the Financial Stability Oversight Council (FSOC) be a mechanism to embrace the congruence principle across institutions and instruments, beyond its current focus on systemically important financial institutions (SIFIs), as it comprises various regulatory representatives and the Secretary of the Treasury of the United States?

The "Hotel California" principle: when regulators support a part of shadow banking ex post by extending the safety net to it, regulators must automatically be bestowed the powers to regulate that part of the financial sector going forward as systemically important. If such a principle had been adopted, it would not have taken as long and with such difficulty to regulate money market funds in the aftermath of the global financial crisis when they were accorded a generous central bank backstop.

Regulatory or macroprudential stress tests should be for the system as a whole rather than just for a set of institutions. That is, stress tests should extend beyond the presently identified SIFIs and provide the analytical basis for whether regulation is congruent across parts of the financial sector or not

(failing which, the congruence principle can be implemented or relevant parts be designated as SIFIs). One example of the success of such an approach might have been the "conversational" FSOC stress test conducted at the time of the European sovereign debt crisis in 2011 when it was clear that the US money market funds were at risk on their commercial paper holdings of European banks and they were persuaded to scale down their exposures in a timely manner.

REFERENCES FOR THE ACHARYA COMMENT

Acharya, Viral V., and T. Sabri Öncü. 2013. "A Proposal for the Resolution of Systemically Important Assets and Liabilities: The Case of the Repo Market." *International Journal of Central Banking* 9, no. 1: 291–350.

Acharya, Viral V., Philipp Schnabl, and Gustavo Suarez. 2013. "Securitization without Risk Transfer." *Journal of Financial Economics* 107, no. 3: 515–36.

COMMENT BY

HYUN SONG SHIN[1] This is an important and timely piece. I would like to draw out one theme in my discussion: the greater heft of nonbank financial intermediaries (NBFIs) in the financial system.

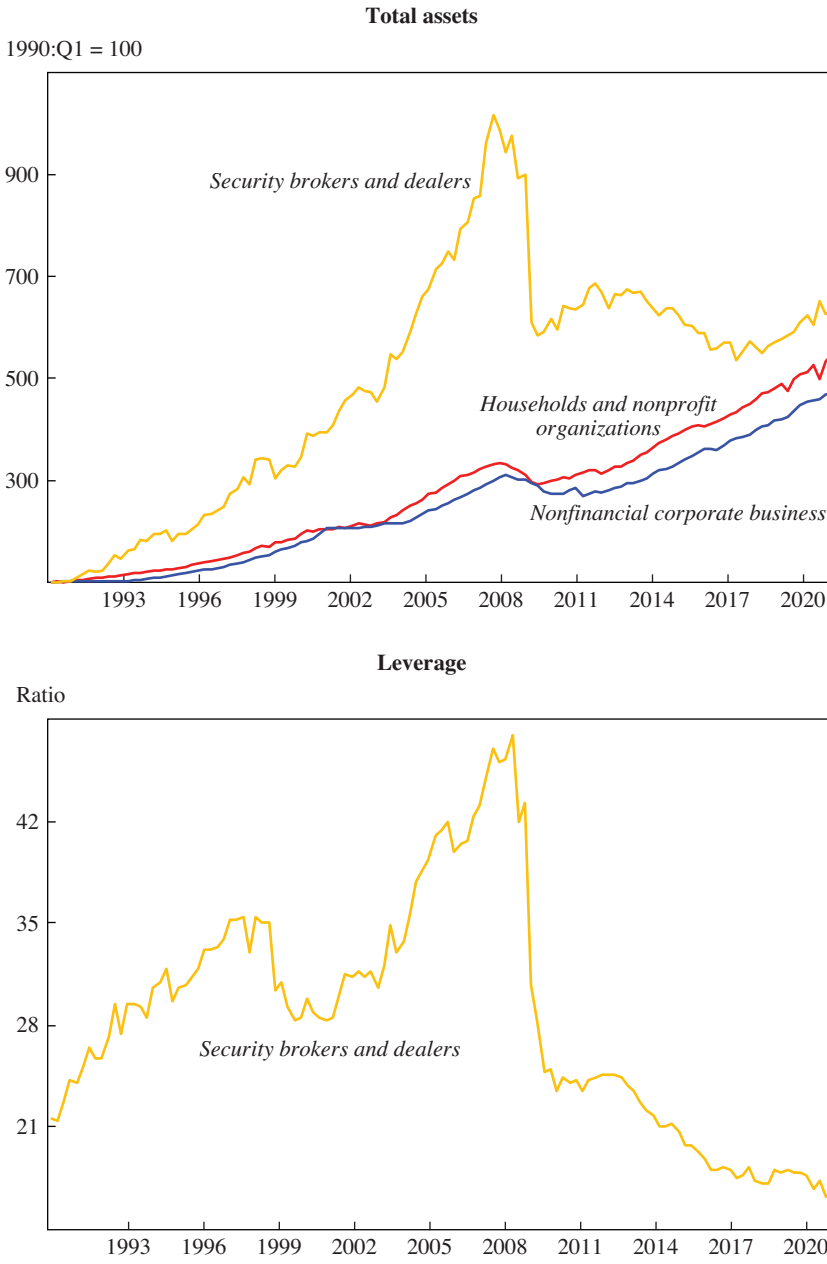
Figure 1 provides a historical sweep of the growth and subsequent contraction of the securities broker-dealer sector in the United States. The top panel shows the time path of total assets of the broker-dealer sector, normalized to 100 in the first quarter of 1990, in comparison to two other sectors—the household sector and nonfinancial corporate sector—also normalized to 100 in 1990:Q1.

On the eve of the financial crisis of 2008, the total assets of the household sector and nonfinancial corporate sector had roughly tripled in size from 1990:Q1, but the broker-dealer sector had grown by about a factor of ten. The immediate precrisis period is discussed in the authors' first example of the growth of mortgage securitizations. The bottom panel of figure 1 on the trajectory of leverage of the broker-dealer sector tells an even more dramatic story. Leverage (defined as total assets divided by book equity) started at just over 20 at the beginning of the period, but rose to around 50 on the eve of the crisis, before dropping sharply with the onset of the crisis.

Thereafter, both the total assets and the leverage of the broker-dealer sector declined further. Total assets of the broker-dealer sector are only modestly higher in relative terms compared to the household and nonfinancial

1. My thanks to Viral Acharya, Sirio Aramonte, Claudio Borio, Stijn Claessens, Neil Esho, Andreas Schrimpf, Vladyslav Sushko, and Nikola Tarashev for helpful discussions and to Giulio Cornelli and Anamaria Illes for research support.

Figure 1. Total Assets and Leverage of the US Securities Broker-Dealer Sector



Sources: Federal Reserve, *Flow of Funds*; BIS calculations.
Note: Leverage (bottom panel) is calculated as total assets divided by equity.

corporate sectors, while leverage has come down to levels that are lower than at any time in recent memory.

When taken at face value, the charts in figure 1 could be read as saying that market-based financial intermediation has been in headlong retreat since the 2008 crisis. However, Metrick and Tarullo's paper tells us that figure 1 is misleading in that respect. While it is true that the on-balance sheet activity of the traditional broker-dealer sector has been subdued, market-based intermediation activity has migrated to places that are not easily captured in the traditional balance sheet aggregates. Figure 1 obscures these structural changes.

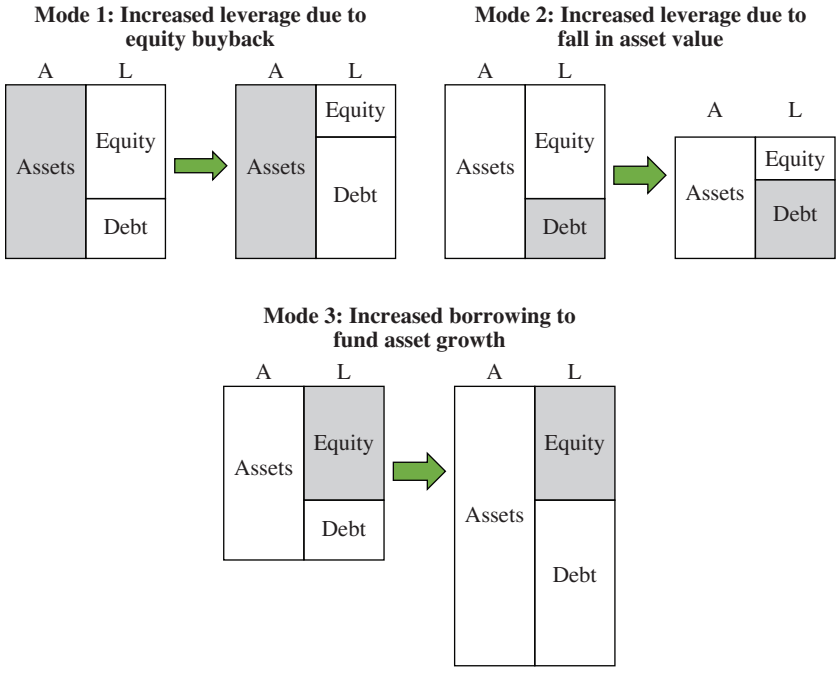
Metrick and Tarullo's discussion of the Treasury market and how it has changed in recent years sheds important further light on these structural changes. They point to the greater role being assumed by hedge funds in both spot and futures markets in Treasury securities. They also highlight the increased importance of central clearing of Treasury repos that has enabled the assembling of leveraged positions by hedge funds by combining long positions in cash Treasuries and short hedging positions in futures exposures. These "mix and match" approaches to assembling an overall position reduce the informativeness of traditional balance sheet series in the Federal Reserve's Flow of Funds as a measure of total exposures.

In the new environment, margin requirements take on a pivotal role for the propagation of financial conditions through the system as a whole. Metrick and Tarullo's main contention is that currently margins are set mainly with a view of the credit risk faced by the clearing house in mind, even when the fluctuations of margin have wider repercussions for the risk-taking capacity of the financial system as a whole. Their notion of congruent regulation is an attempt to formulate a more holistic approach to having in view the risk-taking capacity of the system as a whole. In this context, the leverage ratio of the Basel III bank capital rules assumes an organizing conceptual role.

Metrick and Tarullo's discussion renews attention to the weakness of the traditional picture of the propagation of systemic risk through the "domino" model of cascading defaults. According to the domino model, if Bank A has borrowed from Bank B, while Bank B has borrowed from Bank C, and so on, then a shock to Bank A's assets that leads to default will hit Bank B. If the hit is big enough, Bank B's solvency will be impaired, in which case Bank C would be hit, and so on down the line. Insolvency is seen as the driver of systemic risk in the domino model.

However, while insolvency often figures in systemic crises, it need not do so. Fluctuations in leverage can be a more potent channel of propagation

Figure 2. Three Ways to Increase Leverage



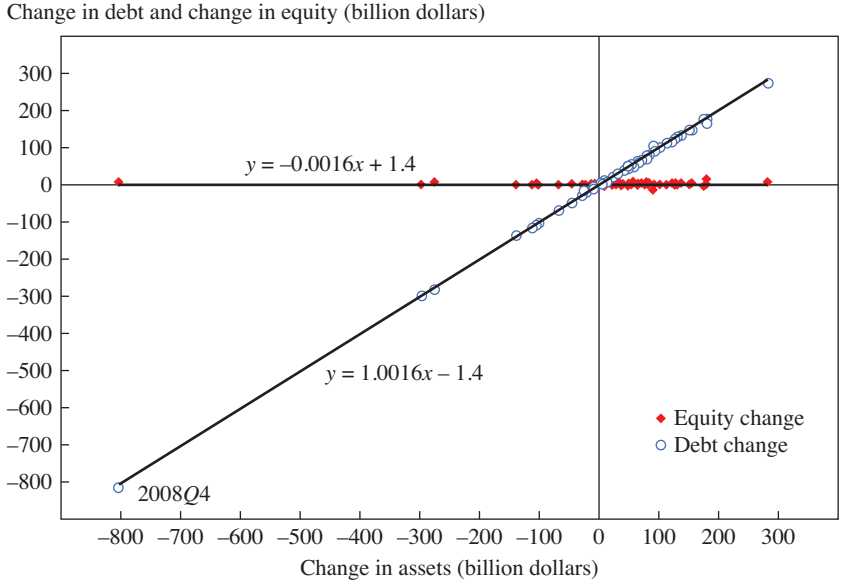
Source: Author.

of systemic risk, especially in settings with market-based intermediation. This is because leverage and balance sheet size move one for one.

To explain, consider three ways of increasing leverage illustrated in figure 2. The first is through an equity buyback through a debt issue (mode 1). The second is through a dividend financed by an asset sale (mode 2). The third is through a reduction in margin requirements that allows the market participant to maintain a larger balance sheet for a fixed amount of its own funds or book equity (mode 3). In each of the three cases, the shaded portion of the balance sheet indicates the component of the balance sheet that is held fixed.

For market-based intermediaries, it turns out that the bottom panel is the relevant case. Leverage moves one for one with asset size, in line with fluctuations in the margin required for each dollar of assets. The one-for-one change in total assets and leverage comes through most clearly for the broker-dealer sector as a whole, as shown in figure 3, taken from Adrian and Shin (2014).

Figure 3. Broker-Dealer Sector Debt Changes and Equity Changes



Source: Adrian and Shin (2014).

In figure 3, the horizontal axis shows the quarterly change in the total assets of the broker-dealer sector from the Federal Reserve's Flow of Funds in dollar terms. The vertical axis then shows how much of the change in assets is reflected in a change in the equity of the sector and how much of the change in assets is reflected in the change in debt. The hollow circles show the relationship between the change in assets and the change in debt, while the diamonds show the relationship between the change in assets and the change in equity. In figure 3, the slope of the relationship between the change in assets and the change in debt is essentially one, meaning that every dollar change in assets goes hand-in-hand with a dollar change in debt. Meanwhile the relationship between the change in assets and change in equity is essentially flat with a slope that is close to zero. This combination of co-movements in balance sheet aggregates is exactly that depicted in the bottom panel of figure 2.

Attainable leverage is the reciprocal of the size of the margin, and so leverage and financing volumes are high in tranquil times but low during stressed times, meaning that financing to others in the system contracts with the onset of stress, sometimes sharply. The dot for 2008:Q4 in figure 3 stands out.

In this way, fluctuations in margin (and the corresponding fluctuations in leverage) are mirrored in the fluctuations in the balance sheet size of system participants and of total financing and degree of interconnections of the system as a whole. In this context, a sharp increase in margins after a protracted period of thin margins will tighten financing conditions for the system as a whole. While insolvencies may exacerbate the stress, they are not necessary for stress propagation. Pecuniary externalities—spillovers that work through prices—can become more potent.

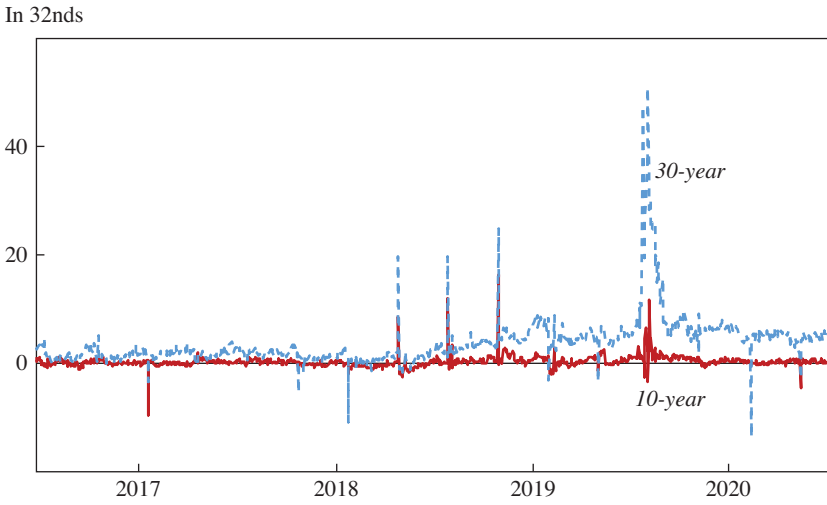
Metrick and Tarullo's second example—their discussion of the Treasury market—underscores these features. The propagation of stress from the fluctuations in leverage should not be viewed in terms of cascading insolvencies of the domino model. Credit risk of the underlying asset is not a necessary condition for stress propagation to emanate from that market. In Morris and Shin (2008), we stressed the point that systemic assets can also be safe assets from a traditional credit risk perspective. Instead, it is the deleveraging channel and the associated pecuniary externalities—the externalities that operate through prices such as spreads and traded risk measures—that can be the most important channel of stress propagation.

Figure 4 illustrates the pecuniary externalities in action during the March 2020 stress episode in the Treasury market. It shows the difference between the price of the notional Treasury securities implied by the respective futures contract (adjusted for the carry that would have come from the coupon of the equivalent cash bond) and the price of the corresponding cheapest-to-deliver cash bond that can be delivered in fulfilment of the futures contract. When the two prices diverge, it means that there is an arbitrage opportunity by taking a long position in one and a short position in the other.

Typically, the futures price–implied Treasury price is higher, reflecting the fact that a futures contract is a zero-money-down bet and does not take up balance sheet capacity at the time when the bet is entered into. In contrast, the equivalent cash bond that is held on the balance sheet will entail a need for balance sheet capacity and associated balance sheet costs. For these reasons, the arbitrage would typically involve taking a short position in the futures contract to hedge the pricing risk of a large, leveraged position in the underlying cash bond.

However, figure 4 shows that this positive spread widened very sharply in March 2020, imposing losses on the convergence trade. The price of the futures-implied Treasury security rose sharply relative to the underlying cash bond. For an arbitrage trader who has a long position in the cash bonds but a short position in the futures-implied bond, this widening would have entailed marked-to-market losses. Schrimpf, Shin, and Sushko (2020)

Figure 4. Price Difference between the Futures-Implied Price of US Treasury Security and the Corresponding Cheapest-to-Deliver Treasury Security



Source: Bloomberg.

Note: Each series is constructed as the difference between the price of the notional bond corresponding to the futures contract adjusted for the cost of carry (coupon income and financing cost) and the corresponding cheapest-to-deliver cash bond.

provided a contemporaneous analysis of the events in the Treasury market in March 2020. Subsequently, more detailed studies that utilize the underlying micro data series have shed further light on the role of leveraged hedge funds in the stress event of March 2020 (Kruttli and others 2021; Barth and Kahn 2021).

Government bond yields provide the benchmark for all other financial assets. Significant disruptions to the functioning of Treasury markets will have broader repercussions, including for the conduct of monetary policy, as we saw again recently in the "taperless tantrum" of late February 2021.

Metrick and Tarullo make the reasonable point that coming to a clear diagnosis of the issue still leaves the question of how the remedies can be put in place when the issues cut across traditional demarcations among regulators.

REFERENCES FOR THE SHIN COMMENT

Adrian, Tobias, and Hyun Song Shin. 2014. "Procyclical Leverage Value-at-Risk." *Review of Financial Studies* 27, no. 2: 373–403.

Barth, Daniel, and Jay Kahn. 2021. "Hedge Funds and the Treasury Cash-Futures Disconnect." Working Paper 21-01. Washington: US Department of the Treasury, Office of Financial Research.

Kruttli, Mathias S., Philip Monin, Lubomir Petrasek, and Sumudu W. Watugala. 2021. "Hedge Fund Treasury Trading and Funding Fragility: Evidence from the COVID-19 Crisis." Working Paper. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econres/feds/hedge-fund-treasury-trading-and-funding-fragility-evidence-from-the-covid-19-crisis.htm.

Morris, Stephen, and Hyun Song Shin. 2008. "Financial Regulation in a System Context." *Brookings Papers on Economic Activity*, Fall, 229–74.

Schrimpf, Andreas, Hyun Song Shin, and Vladyslav Sushko. 2020. "Leverage and Margin Spirals in Fixed Income Markets during the Covid-19 Crisis." *BIS Bulletin* no. 2. Basel: Bank for International Settlements.

GENERAL DISCUSSION Don Kohn commended the authors for doing a great job defining the central problem of risk moving away from the banking system, the regulatory arbitrage, and financial stability risk. He concurred with the congruence concept of linking up regulations to achieve same leverage across transactions. Kohn appreciated the authors' cleverness in finding existing authority for proactive coordination across agencies.

However, Kohn expressed skepticism at agency willingness for buying into these recommendations. Recalling resistance to what seemed like simple money market fund reform five years ago, he observed that it might be very hard to achieve this coordination.[1] As a potential legislative solution for implementation, Kohn suggested that agencies on the Financial Stability Oversight Council (FSOC) should be given a financial stability mandate in addition to their existing mandates. Additionally, they must be made to create a financial stability office that engages with all the rule making and interacts with other offices. Kohn expressed doubt at agencies taking actions recommended by the authors without a financial stability mandate.

Austan Goolsbee pointed out that legislative and political figures will work to prevent the proposed regulation, as reflected by the existence of the shadow banking system. Subsequently, Goolsbee wondered if there is a way to quantify how much more damaging it is to have the Federal Reserve do emergency rescues or interventions instead of overturning the politics of the regulatory system.

Daniel Tarullo responded by first addressing the political economy considerations of Kohn's and Goolsbee's questions. He agreed with Kohn and

1. Money Market Fund Reform; Amendments to Form PF, 79 Fed. Reg. 47,735 (August 14, 2014).

Goolsbee in anticipating the institutional and political difficulties tied to implementation. Addressing Kohn, Tarullo argued that, from a lawyer's perspective, the Dodd-Frank Act already incorporates financial stability into the SEC's mandates, although that is not explicitly stated. Tarullo remarked that legislation reaffirming this mandate would be great. He speculated that President Biden's appointees to the regulatory financial agencies would have greater inclination to take on these reforms as compared to their immediate predecessors or even compared to President Obama's appointees.

With respect to Goolsbee's quantification remark, Tarullo opted to give a qualitative response. In the case of lack of regulatory prospects leading to a free ride for nonbank financial intermediaries, Tarullo said that he would anticipate an accelerated outmigration of financial activity from the prudentially regulated sector, which would start eroding the franchise value of the existing prudentially regulated structure. While he conceded that the paper's recommendations are somewhat messy, he expressed skepticism at the prospects of other alternatives. He agreed that some of Viral Acharya's recommendations would underscore responsibility even short of Kohn's legislative approach. However, he argued that if the Treasury and the Federal Reserve resist the kind of regulation he and Andrew Metrick have recommended, then the Federal Reserve would have to make decisions about providing liquidity every time there's a market dislocation and weigh the moral hazard cost of reinforcing market reliance on the Federal Reserve during distress.

Further addressing Acharya's suggestions about using other institutional mechanisms to implement the congruence principle, such as FSOC, extending the safety net, and system-wide stress tests, Tarullo commented that one of the mechanisms was partially incorporated into the Dodd-Frank Act: even if institutions like Morgan Stanley and Goldman Sachs were to divest their depository institutions and cease to be bank holding companies, they would continue to remain regulated by the Federal Reserve because they received government capital during the global financial crisis. But, he also stated he would be in favor of generalizing this requirement in future legislation. Tarullo remarked that President Obama made similar suggestions in his 2008 speech as a candidate at the Cooper Union by proposing that any entity getting assistance from the Federal Reserve during a financial crisis should by definition be a prudentially regulated entity.[2] Tarullo observed that the Federal Reserve injected liquidity into the markets rather than

2. Barack Obama, "Renewing the American Economy," speech given at the Cooper Union, New York, March 27, 2008, https://www.nytimes.com/2008/03/27/us/politics/27text-obama.html.

financial institutions during the great financial crisis, which he speculated would make it harder to implement this idea.

With regards to stress testing, Tarullo reflected that, while it's a good idea, system-wide stress testing would be tough to implement from a bureaucratic standpoint, especially in the balkanized regulatory system in the United States. He contemplated that its implementation would relatively be easier in the United Kingdom because of its more unified regulatory approach. Further, he acknowledged that from a policy and analytical standpoint, stress testing is unimpeachably a good idea. Finally, Tarullo agreed with the general comments on FSOC and further argued that the paper's proposal for the Treasury to take leadership on Treasury-backed repo would aid the Treasury to leverage its position within the FSOC.

Metrick observed that many comments questioned the practicality of the paper's suggestions. In response, he noted that the paper's proposed concept of congruence seeks to address these concerns by leveraging existing statutory authority. He argued that congruence is different than regulating function rather than form, which would indeed require legislation and a paradigm shift in the United States. Instead, he continued, the congruence concept encourages agencies to consider their existing authority in the context of systemic risk. For example, instead of looking at its margining authority in context of protecting investors, the Securities and Exchange Commission should consider it as being congruent to the financial stability concerns of bank regulators. Metrick concluded that this is what makes the congruence principle different than regulating form and function and achievable under current statutory authority.

Andrew Atkeson noted that Larry Fink had stated in a prior conversation with him that he believed the Employee Retirement Income Security Act (ERISA) was responsible for Wall Street getting so big. Atkeson reflected that this might be the case because the funding of defined benefit pension funds required under ERISA created big pools of institutional money. He observed that this is a root cause of the demand for risk-taking in the financial sector because the taxpayers would bail out these pension funds if they go bad. Accordingly, Atkeson wondered if instead of regulating intermediaries, it would make sense to directly regulate these funds and treat them as investors in the United States and globally.

Metrick responded by stating that while he didn't think these funds have historically been a part of the problem, they could be a part of the solution. Atkeson argued that the funds could indirectly have been a problem through hedge funds, which Metrick agreed was an interesting thought. Tarullo also agreed that it was interesting but noted that he would be reluctant to identify

that as today's source of motivation for many things, which would shift to different channels tomorrow. Tarullo argued that financial stability regulation should focus on sources of financial risk. However, he acknowledged that intermediaries such as mutual funds and exchange traded funds that don't take on risk themselves but are maturity transformers might need direct attention because the scope of their maturity transformation has grown exponentially since the global financial crisis.

Acharya observed that while it is possible that the pension funds' search for yield might manifest itself through hedge funds or other institutional investors, he wondered if the returns are comparable to traditional fixed-income investments. He highlighted this as an important financial stability angle to monetary policy, which needs to distinguish between longer-term value enhancing investments and longer-term high-risk speculative investments. Acharya speculated that the root cause of the problem is that the long end of traditional fixed-income investments and other safe assets does not yield enough returns for pension funds, which should potentially be tackled by monetary policy rather than financial regulation.

Panel on Fifty Years of BPEA*'s Contributions to Macroeconomics and Policy*

Editors' Note

Brookings Papers on Economic Activity (BPEA) marked its fiftieth anniversary in 2020. Papers by three longtime contributors highlighted *BPEA's* seminal research over the years in areas at the heart of macroeconomic policymaking: labor markets, productivity and growth, and monetary policy. Robert E. Hall and Robert J. Gordon participated in the first *BPEA* conference in April 1970, and Alan S. Blinder was a participant in the Fall 1972 conference. All three had fresh PhDs from the Massachusetts Institute of Technology when their service on the panel began. Recordings of their retrospective presentations can be found on the Brookings website at https://www.brookings.edu/events/bpea-spring-2021-conference/.

ROBERT E. HALL
Stanford University

The Brookings Panel's Contributions to Research on Labor Markets

ABSTRACT On the occasion of the fiftieth anniversary of the Brookings Panel on Economic Activity, I review the extensive body of research that has appeared in the *Brookings Papers on Economic Activity* (*BPEA*) on the labor market. Much of the research deals with unemployment, a topic of great interest in macroeconomic analysis and policy. I trace the evolution of modern economic analysis of unemployment and the major contributions relating to unemployment in the pages of the *Brookings Papers*. I also review a number of important contributions to other aspects of labor economics that are part of the *BPEA* legacy.

The Brookings Panel on Economic Activity, and its journal, the *Brookings Papers on Economic Activity* (*BPEA*), has played a key role in the evolution of scientific understanding of the US labor market over the past fifty years. As in other branches of macroeconomics and related specialties, the Brookings Panel has developed a unique position in the research process and in the dissemination of research findings at the intersection of labor economics and macroeconomics. Major new ideas and theoretical constructs have informed the panel's research and papers, always accompanied by careful use of the relevant data. Although the Brookings Panel has remained faithful to its founding goal of sponsoring research and publishing papers that informed current policy debates, it has also established a leading position in basic research on labor market issues. Ever since the price of a room at the old Dupont Plaza hotel was $14, it has been my privilege to be involved in the panel's activities as an occasional author, frequent discussant, and inveterate formulator of off-the-wall remarks from the floor.

Conflict of Interest Disclosure: The author did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this paper. He is currently not an officer, director, or board member of any organization with an interest in this paper.

My remarks come in two parts. First, the labor market topic that has received the most attention from the Brookings Panel is unemployment. This preoccupation is not surprising. Unemployment is a key indicator of the state of the aggregate economy. Monetary policy and national fiscal policy aim to stabilize the unemployment rate at a low level, possibly as low as 3.5 percent, the rate prevailing before the roof fell in last year. Unemployment receives as much attention from macroeconomists as from labor economists. Modeling of unemployment advanced enormously over the past fifty years, and the advances were the subject of numerous well known and heavily cited Brookings papers.

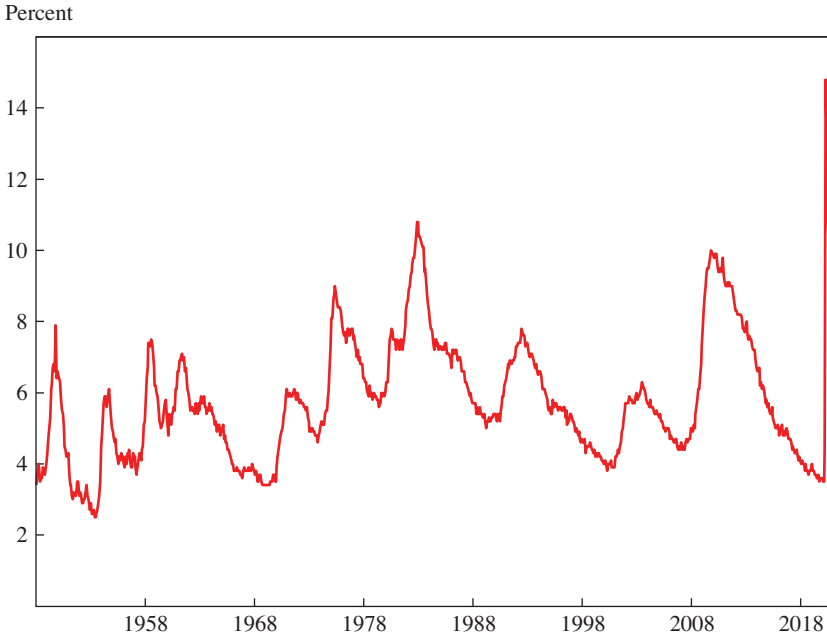
Second, Brookings papers on labor market topics apart from unemployment have also had high impact. I will note the contributions of some specific papers in four areas: labor dynamics across the United States; effects of rising immigration and international trade on wages; the decline in labor's share of national income; and rising volatility of individual earnings over time.

I. Unemployment

Figure 1 shows the unemployment rate over the period that it has been measured scientifically and consistently in a survey of a large number of US households. Unemployment is the quintessential cyclical measure—there is no need for gray bars in the figure to identify recessions. Unemployment has no trend. It began around 3 percent in 1948 and stood in February 2020 near the same level. In the 1970s and 1980s, unemployment was generally higher. In particular, unemployment declined to only 5 or 6 percent, compared to 3 or 4 percent in the earlier and later decades. Unemployment leaps upward in every recession, most notably in the onset of the pandemic in the spring of 2020.

The traditional macroeconomic view of unemployment was simply the difference between labor supply and labor demand. At the personal level, being unemployed was a state that a low percentage of workers occupied in normal times but which doubled in recessions and increased way more in the pandemic. In 1970, coincidentally the year of the Brookings Panel's founding, a book appeared that came to be known as the "Phelps volume" (Phelps 1970). Its editor, Edmund Phelps, had spotted a new development in economic theory promising an alternative to Walrasian competitive equilibrium and to the idea that disequilibrium could be modeled as the difference between Walrasian supply and demand. In the concluding chapter in the volume, Phelps, along with Sidney Winter, wrote: "A landing on the

Figure 1. US Unemployment Rate Starting in 1948



Source: Current Population Survey.

non-Walrasian continent has been made. Whatever further exploration may reveal, it has been a mind-expanding trip: We need never go back to $\dot{p} = \alpha(D - S)$ and $q = \min(D, S)$" (337).

I was a fellow traveler with the authors of the volume but not an author myself, and I was the (highly sympathetic) reviewer of the volume for the *Journal of Economic Literature* (Hall 1972).

Phelps and Winter proved right on the first point—the landing on that continent has been as successful as the British settlement of North America. However, they were wrong on the second. The branch of macro that deals with the output gap and the Phillips curve—Europe, to continue the continental analogy—remains equally successful, though outside the scope of my remarks.

The search-and-matching model that germinated in the Phelps volume started from the proposition that unemployment was a purposeful activity of people who desired to work and were in the process of searching for work. To explain the continuing presence of unemployment, the new

model invoked a steady flow of job losers and job leavers. Unemployment represented the stochastic equilibrium of inflows of workers to the pool of job seekers and outflows from the pool.

The search-and-matching model grew in influence to the point of a Nobel Prize in 2010, awarded to Peter Diamond, Dale Mortensen (author of a chapter in the Phelps volume), and Christopher Pissarides. Many established macroeconomists in 1970 gave it a chilly reception, arguing that search theory was blaming the victim by making job search somehow voluntary. That condemnation has gradually declined. Most macroeconomists today view unemployment as the result of rational conduct of job seekers in the face of labor market frictions that impede the job-finding process.

We now call the search-and-matching model the DMP model. Perhaps the single most important contribution of the DMP model to economic theory is to make rigorous theoretical sense out of the concept of labor market tightness. The concept was obviously important in practice, but it had no previous counterpart in theory. Sometimes, notably in the years just before the pandemic, the market is tight. Job seekers find jobs quickly, and employers have to wait to find qualified applicants for jobs. In other times, such as 2010, the reverse holds—jobs are hard to find, and vacancies are easy to fill. The DMP model defines tightness as the ratio of vacancies to unemployment.

The DMP model has two components. The first, search and matching, describes the frictional job-finding process. Its centerpiece is the matching function. In a given labor market, the volume of job seekers and the volume of job openings are factors of production that combine, as in a production function, to generate a flow of newly filled jobs. The matching function determines two key flow rates as functions of tightness, the job-finding rate for job seekers (the ratio of jobs filled to the number of job seekers) and the job-filling rate for employers (the ratio of jobs filled to the number of vacancies). The matching function also lies behind the Beveridge curve, which traces out the variations in tightness in the unemployment-vacancy space—vacancies are high when unemployment is low in a tight market and unemployment is high and vacancies low in a slack market.

The second component of the DMP model involves wage determination. The incentive to create a job is provided by the job value, which is the present value of the difference between a worker's contribution to revenue and the worker's wage over the duration of the new job. In a frictional labor market, where employers exert effort to get in touch with prospective workers, the incentive will be present in equilibrium—wages will be below productivity. A key assumption of the DMP model is that employers

maintain vacancies at the level where the incremental vacancy has zero net contribution to profit.

The two components of the DMP model connect because the zero-profit condition determines tightness, tightness determines the job-finding rate, and the job-finding rate determines the path of unemployment. From an initial point after an adverse shock that has caused a jump in unemployment, but with normal conditions gradually raising tightness and thus the job-finding rate during the recovery, applying that rate to higher unemployment causes the excess unemployment to disappear over time, thanks to the rising flow of workers into jobs.

What kind of a shock would cause an increase in unemployment in a recession? Only a decline in the job value. In the original developments of the DMP model, that decline took the form of a drop in productivity. Whereas traditional thinking involved a drop in aggregate demand, in the DMP model the source had to be something more specific. Though productivity is an obvious source as a matter of theory, measured productivity fluctuations do not constitute a plausible candidate in data for the past fifty years, especially recently. Recent work has turned to rises in the discount rate as a source of sharp declines in the job value.

Many Brookings papers dealt with the issues of unemployment and the DMP model. The first issue was the new view of unemployment as a frictional process that involved more than a simple gap. My paper (Hall 1970) appeared in the first year of the panel's existence and the year of publication of the Phelps volume. My Brookings paper was the first of quite a few in *BPEA* to study flows into and out of unemployment for various demographic groups. It concluded: "Unemployment is high at full employment both because (1) normal unemployment remains high—the natural flow of workers through the labor market is high; and (2) there is an additional component of abnormal unemployment—members of some groups in the labor force do not follow definite careers but change frequently and erratically from one job to another, experiencing unemployment with most changes" (372–73).

Perry (1972) was the next in the line of Brookings papers that studied labor market flows in the Current Population Survey. Perry extended the investigation into what the literature now calls the three-state model, adding consideration of time spent out of the labor force to the story. Charles Holt's discussion nicely summarized its conclusion: "He shows clearly and dramatically the dynamic character of most unemployment. For most workers, unemployment is a state through which many pass, rather than . . . a condition that constitutes a chronic problem for a fixed

group of workers" (Perry 1972, 282). Holt's commentary also included a Cobb-Douglas matching function in exactly the same functional form and notation that has been used in hundreds of search-and-matching papers in subsequent decades (286).

Next in order in this line of Brookings papers is Marston (1976). This paper continued the development of the three-state dynamic model. Marston introduced the study of what has come to be called the "ins and outs" of unemployment—rates of inflow to unemployment from job loss and rates of outflow from job finding. He also documented that the inflow also included entry to unemployment of people previously out of the labor force and exits from unemployment to out of the labor force.

Clark and Summers (1979) detected a conclusion that economists were drawing from the findings of the studies of flows into and out of unemployment. The conclusion was that unemployment arose from short spells of job seeking among a broad swath of the labor force, rather than a concentration of extensive unemployment among a small fraction of the population. Clark and Summers showed that this conclusion was incorrect; even though most spells of unemployment are short, most unemployment occurs among people suffering repeated long spells of unemployment.

Summers (1986) tackled the explanation of the bulge of unemployment that is apparent in figure 1 after 1970. The expansion of the 1960s drove the rate down to 3.4 percent. The first expansion of the 1970s got to 4.6 percent and the second expansion to only 5.6 percent. Summers was writing partway through the expansion following the deep recession of the early 1980s, which had propelled the rate to its all-time high of 10.8 percent. He recognized that part of the rise occurred because of demographic shifts toward groups with higher normal unemployment, but he concluded that there was more to the story: "increases in unemployment are a serious problem because they are concentrated among mature men, job losers, and the long-term unemployed" (340). The expansion underway when Summers was writing got unemployment down to 5.0 percent, the strong expansion of the 1990s achieved 3.8 percent, the weaker expansion of the 2000s reached 4.4 percent, and the vigorous expansion of the 2010s got back to 3.5 percent. So the changes that Summers noted were generally reversed. At least some of his observations—such as the importance of job losers and the elevation of long-term unemployment—were the lingering result of the bad recession four years earlier. Similar issues came up during the earlier years of the expansion that ended in the spring of 2020.

Abraham (1987) followed Summers's paper by a year and considered the same apparent systematic rise in unemployment from 1970 to

the mid-1980s. She studied the issue from the perspective of the Beveridge curve, which required her to deal with the absence of direct measures of vacancies in the United States. In 1987, this meant validating a measure based on the volume of help-wanted advertising. She diagnosed an outward shift of the Beveridge curve, which accounts for higher average unemployment. Later data have confirmed the continuation of that shift but also found it to be offset by a downward trend in the flow into unemployment, which is consistent with the restoration of earlier average levels of unemployment in the 1990s through the present. During her long term as commissioner of the Bureau of Labor Statistics, starting in 1993, she launched the Job Openings and Labor Turnover Survey (JOLTS), which has provided reliable, economy-wide data on vacancies since late 2000.

Blanchard and Diamond (1989) investigated the labor market turnover process in the framework of the Beveridge curve. A year earlier, Pissarides (1988) had published a related paper that laid out what became the theory of labor market tightness, the matching component of the DMP model, in Mortensen and Pissarides (1994). Blanchard and Diamond were mainly concerned with separating the effects of shocks that moved the market along its Beveridge curve and shocks that moved the Beveridge curve. They observe: "Aggregate activity shocks drive unemployment and vacancies in opposite directions, causing counterclockwise movements around a downward-sloping locus in the Beveridge space. Reallocation shocks lead instead to movements along an upward-sloping locus, to parallel movements in unemployment and vacancies" (2–3). Their aggregate activity shocks operate through changes in market tightness in the subsequent literature, though they do not invoke that concept. They found, in concert with a voluminous later literature, that the disproportional source of the volatility of unemployment and vacancies was the aggregate activity shock.

Blanchard and Diamond (1990) followed up their paper the year before by studying two bodies of data in parallel, firm-level employment changes and household-level flows among employment, unemployment, and nonmarket pursuits. They showed that recessions, which involve contractions in total employment, achieve most of the decline from major cuts in some firms, rather than from diminished flows of hires. In most theoretical treatments of hiring and firing decisions, the tendency for new hires to have lower surpluses than incumbents should make the hiring rate more sensitive to adverse events than the firing rate. With respect to the cyclical flows at the household level, they find important changes over the cycle in flows into and out of the labor market.

Krueger, Cramer, and Cho (2014) made a key contribution to research in labor market dynamics by their critique of the prevailing practice of studying monthly transition rates as descriptions of behavior. They found that the probability that a job seeker finds a job in the coming month is not really indicative of job-finding success—job-finding rates seriously overstate the probability of finding a stable job. In data from the Current Population Survey, it is a good idea to look at what happens over the entire sixteen-month span that the survey covers for each respondent. The paper applies this insight to develop a picture of the experiences of job seekers who have been looking for many months. In the aftermath of serious recessions, notably the one that started in late 2007, substantial populations of these long-term unemployed accumulate.

Davis and von Wachter (2011) studied administrative data on workers who lost jobs where they had substantial tenure. They followed job losers for many years as they recovered from the adverse effects of job loss. This line of research is the ultimate application of the idea of tracking individuals in panel data rather than trying to infer experiences by estimating a model based on one or a few transitions. Earlier research of this type used panel surveys, but administrative data are more plentiful and accurate. The results show that displacement from a job that has proven durable results, typically, in a year or two of depressed employment—presumably a number of spells of unemployment or time out of the labor force—followed by lower wage rates for many years, often until retirement. Models in the DMP framework are only beginning to absorb the teachings of this literature, where the paper by Davis and von Wachter is prominent.

Gordon (1973) quantified the gains and losses from reductions in unemployment. He provided a comprehensive analysis of the differences in the economy between 4 and 5 percent unemployment. From the starting point that a small perturbation from an optimum in a friction-free economy has no effect on welfare, Gordon studied, in incredible detail, what we would now call wedges. The obvious wedge is the income tax, but the paper considers many others. His conclusion was that wedges are sufficiently big that the naive analysis based on Okun's law—real GDP would rise by 2.7 percent of real GDP for each decline in unemployment of one percentage point—is only exaggerated by 0.4 percentage points. If this number is correct, analyses in the more recent literature on the burden of wedges have seriously underestimated that burden.

Okun's (1973) paper on the benefits of a low-unemployment economy appeared in the same issue of *BPEA* as Gordon's paper quantifying those

benefits. He noted that "unemployment [is] merely the tip of the iceberg that forms in a cold economy" (208). In addition to adjustments that are also present in Gordon's paper, Okun focused on employee upgrading.

Katz and Krueger (1999) present a retrospective discussion of the tight labor market of the 1990s, written at a time when labor market conditions were similar to those of 2019. The paper's main contact with the ideas considered here is its documentation of an inward shift of the Beveridge curve, just before the onset of reliable data on vacancies from JOLTS. The authors were too cautious to quantify the shift in terms of the shift in unemployment conditional on the vacancy rate, but their figure 5 easily supports a 2 percentage point drop in tightness-adjusted unemployment. They reported a 0.4 percentage point decline attributable to the aging of the labor force and a 0.2 percentage point decline from rising incarceration.

Aaronson and others (2006) provide a comprehensive review of data on labor force participation. The participation rate is only barely cyclical—when employment plunges in a recession, unemployment rises by close to the same amount. The DMP canon, Mortensen and Pissarides (1994), studied a population who all participated. The transition to the three-state DMP-style model in recent years has made the labor force participation rate an important topic in that body of thought. Aaronson and others did not enter the territory of modeling labor supply, but they provided an informative account of the participation rate over time, broken down into detailed demographic groups. Prior to 2000, overall participation rose because rising rates among women offset gently declining rates among men. But the rise among women ended in that year, and overall participation has been declining, with only a small reversal recently. The equations fitted by Aaronson and others permitted calculations of the likely path of participation in future years. The paper forecasted a decline of 3.6 percentage points from 2005 to 2015. The forecasts of the Congressional Budget Office, the Bureau of Labor Statistics, and the Social Security Administration were for declines of 1.5 percentage points or less (Aaronson and others 2006, table 6).

Shortly after the financial crisis in 2008, an unusual decline in participation occurred. Many observers thought that the extremely slack labor market in 2009 and 2010 had changed the earlier rule that the increase in unemployment roughly equaled the decline in employment. There was concern that the rise in unemployment therefore understated the effect of the recession in the labor market. Aaronson and others (2014) tackled this

issue. Their conclusion was that the decline in participation was not only foreseeable but had been foreseen in Aaronson and others (2006). Their figure 1 shows that the actual participation rate through 2014 tracked the forecast from the 2006 paper remarkably well. Because the 2006 paper made no adjustment for the unforeseeable tragedy of the financial crisis, the conclusion appears to be that even a large and persistent increase in unemployment has little effect on participation.

Abraham, Haltiwanger, and Rendell (2020) studied the seemingly super-tight labor market of 2019 in a framework that recognized heterogeneity among the unemployed and other job seekers. After adjustment for differences in base-year job-finding rates, the authors show that some anomalies that appeared in the data, notably in a large favorable shift of the Beveridge curve, were artifacts that disappeared in the adjusted data. They concluded that the apparatus underlying the DMP model performs reliably even at unemployment rates under 4 percent once composition effects are taken into account.

Last but conspicuously not least was the explosion of unemployment in April 2020 from the pandemic. As the figure shows, unemployment reached much higher levels than in any other time in the figure. The Brookings Panel swung into immediate action, scheduling a special meeting on the macroeconomics of the pandemic in June 2020. Two papers focused on the labor market. Cajner and others (2020) documented the huge 21 percent decline in employment that occurred in late March and April, as the economy shut down. It also showed that recalls of workers on temporary layoff occurred in May, anticipating the sudden importance of the layoff-recall process that dominated labor market dynamics in later months of the pandemic. Bartik and others (2020) studied a variety of high-frequency data sources to demonstrate the concentration of reduced hours of work in the retail and hospitality sectors. The authors found that the vast majority of laid-off workers expected to be recalled, and some had already been recalled by June.

At the regular September 2020 meeting of the panel, Gallant and others (2020) presented a detailed structural model of the pandemic labor market. They emphasize the importance of treating unemployed workers who expect to be recalled differently from those who have definitively lost jobs. For the first time, people holding jobs but not working or being paid by their employers were an important fraction of the unemployed. The monthly probability of recall and resumption of work is much higher for those on layoff than for those suffering job loss. Figure 1 shows that unemployment fell much faster from its peak in April 2020 than it ever had in the aftermath

of earlier spikes in unemployment, a strong confirmation of the altered structure of the labor market diagnosed in the paper.

II. Other Labor Topics

BPEA has published dozens of papers in the macro-labor subject area that are not focused specifically on unemployment. Five stand out in my memory.

Blanchard and Katz (1992) studied the state-level dynamics of employment and unemployment, reaching two famous conclusions: "a state typically returns to normal after an adverse shock not because employment picks up, but because workers leave the state" and "in response to an adverse shock in employment, nominal wages decline strongly before returning to normal after approximately 10 years. This decline triggers some recovery in employment, but the response of job creation to wage declines is not sufficient to fully offset the initial shock" (3).

Borjas, Freeman, and Katz (1997) asked the perennial questions: Are low-skill domestic workers harmed by immigration? And, are workers harmed by international trade? The answers were yes and no. The effects of immigration operate across all sectors; many low-skill immigrants work in construction and services. The effects are large and geographically concentrated. Immigration is concentrated in high-skill individuals and in those lacking even high-school level education. The effects of trade operate only through tradable goods and are small.

Two important papers have tackled the issue of the decline of labor's share of national income in recent decades. First was Elsby, Hobijn, and Şahin (2013). Their paper disposed of two explanations, finding little role for capital-labor substitution and for the decline of unionization. They pointed out that the treatment of the self-employed overstates the decline and finds that the decline is mostly confined to manufacturing. In that connection, they pointed out that the offshoring of the labor-intensive components of supply chains is a growing source of decline in the measured labor share.

Rognlie (2015) built on the earlier work of Elsby, Hobijn, and Şahin (2013), taking particular aim at the suggestion of Piketty and others that capital accumulation was the driver of the decline in the labor share. Rognlie observed that the comprehensive measure of the share used in that literature includes housing, which accounts for a large part of the measured decline. He favored studying the corporate sector, in part to avoid the problem of measuring the share for the self-employed. He argued for using capital income net of depreciation in calculating the labor share.

Gottschalk and Moffitt (1994) considered the body of research as of the mid-1990s that demonstrated high and rising dispersion of labor earnings across individual workers. That research had interpreted the dispersion as reflecting inequality. This paper made the key point that measured cross-sectional dispersion combined inequality in the permanent component of wages with the variability of earnings over time at the individual level. In the subsequent twenty-five years, the availability of large panels of administrative data on earnings, and the computing power to study those data, has allowed researchers to follow up the authors' insight and make big advances in our understanding of individual earnings dispersion. For example, Fatih Guvenen has access to every single W-2 form filed with the IRS from 1978 to recent years.

III. Concluding Remarks

The story of the founding of the Brookings Panel is almost lost in the mists of time. Prior to the panel's founding in 1970, the Brookings Institution's main involvement in macroeconomics was the sponsorship and funding of the Brookings model of the US economy. Though that model captured the attention of macroeconomists in its time, the late 1960s, today it is forgotten, and its vestiges remain only in the form of models used in central banks. A joke circulated in those days that the main purpose of big models was to see that all accounting identities were satisfied in the model's output.

The panel's founders, Arthur Okun and George Perry, believed that macro would advance as a science if we attacked the subject in chunks, issue by issue and event by event. The founding motto was "NO big models," and none ever appeared in the pages of *BPEA*. I hardly need to say how much I agreed with the motto, then and now.

The Brookings Panel was a complete innovation. Nothing like it existed, in macro or in any other branch of economics. I congratulate the designers for a remarkable and durable achievement.

References

Among the papers cited here, I served as a discussant for Perry (1972), Clark and Summers (1979), Blanchard and Diamond (1989, 1990), Blanchard and Katz (1992), Davis and von Wachter (2011), and Aaronson and others (2014).

Aaronson, Stephanie, Tomaz Cajner, Bruce Fallick, Felix Galbis-Reig, Christopher Smith, and William Wascher. 2014. "Labor Force Participation: Recent Developments and Future Prospects." *Brookings Papers on Economic Activity*, Fall, 197–255.

Aaronson, Stephanie, Bruce Fallick, Andrew Figura, Jonathan Pingle, and William Wascher. 2006. "The Recent Decline in the Labor Force Participation Rate and Its Implications for Potential Labor Supply." *Brookings Papers on Economic Activity*, no. 1: 69–134.

Abraham, Katharine G. 1987. "Help-Wanted Advertising, Job Vacancies, and Unemployment." *Brookings Papers on Economic Activity*, no. 1: 207–43.

Abraham, Katharine G., John C. Haltiwanger, and Lea E. Rendell. 2020. "How Tight Is the US Labor Market?" *Brookings Papers on Economic Activity*, Spring, 97–138.

Bartik, Alexander W., Marianne Bertrand, Feng Ling, Jesse Rothstein, and Matthew Unrath. 2020. "Measuring the Labor Market at the Onset of the COVID-19 Crisis." *Brookings Papers on Economic Activity*, Summer, 239–68.

Blanchard, Olivier Jean, and Peter Diamond. 1989. "The Beveridge Curve." *Brookings Papers on Economic Activity*, no. 1: 1–76.

Blanchard, Olivier Jean, and Peter Diamond. 1990. "The Cyclical Behavior of the Gross Flows of U.S. Workers." *Brookings Papers on Economic Activity*, no. 2: 85–155.

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity*, no. 1: 1–75.

Borjas, George J., Richard B. Freeman, and Lawrence F. Katz. 1997. "How Much Do Immigration and Trade Affect Labor Market Outcomes?" *Brookings Papers on Economic Activity*, no. 1: 1–90.

Cajner, Tomaz, Leland D. Crane, Ryan A. Decker, John Grigsby, Adrian Hamins-Puertolas, Erik Hurst, Christopher Kurz, and Ahu Yildirmaz. 2020. "The US Labor Market during the Beginning of the Pandemic Recession." *Brookings Papers on Economic Activity*, Summer, 3–33.

Clark, Kim B., and Lawrence H. Summers. 1979. "Labor Market Dynamics and Unemployment: A Reconsideration." *Brookings Papers on Economic Activity*, no. 1: 13–72.

Davis, Steven J., and Till von Wachter. 2011. "Recessions and the Costs of Job Loss." *Brookings Papers on Economic Activity*, Fall, 1–72.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2013. "The Decline of the U.S. Labor Share." *Brookings Papers on Economic Activity*, Fall, 1–52.

Gallant, Jessica, Kory Kroft, Fabian Lange, and Matthew J. Notowidigdo. 2020. "Temporary Unemployment and Labor Market Dynamics during the COVID-19 Recession." *Brookings Papers on Economic Activity*, Fall, 167–216.

Gordon, Robert J. 1973. "The Welfare Cost of Higher Unemployment." *Brookings Papers on Economic Activity*, no. 1: 133–205.

Gottschalk, Peter, and Robert Moffitt. 1994. "The Growth of Earnings Instability in the U.S. Labor Market." *Brookings Papers on Economic Activity*, no. 2: 217–72.

Hall, Robert E. 1970. "Why Is the Unemployment Rate So High at Full Employment?" *Brookings Papers on Economic Activity*, no. 3: 369–410.

Hall, Robert E. 1972. Review of *Microeconomic Foundations of Employment and Inflation Theory*, by Edmund S. Phelps, ed. *Journal of Economic Literature* 10, no. 1 (March): 65–66.

Katz, Lawrence F., and Alan B. Krueger. 1999. "The High-Pressure U.S. Labor Market of the 1990s." *Brookings Papers on Economic Activity*, no. 1: 1–87.

Krueger, Alan B., Judd Cramer, and David Cho. 2014. "Are the Long-Term Unemployed on the Margins of the Labor Market?" *Brookings Papers on Economic Activity*, Spring, 229–80.

Marston, Stephen T. 1976. "Employment Instability and High Unemployment Rates." *Brookings Papers on Economic Activity*, no. 1: 169–203.

Mortensen, Dale T., and Christopher A. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies* 61, no. 3: 397–415.

Okun, Arthur M. 1973. "Upward Mobility in a High-Pressure Economy." *Brookings Papers on Economic Activity*, no. 1: 207–52.

Perry, George L. 1972. "Unemployment Flows in the U.S. Labor Market." *Brookings Papers on Economic Activity*, no. 2: 245–92.

Phelps, Edmund S., ed. 1970. *Microeconomic Foundations of Employment and Inflation Theory*. New York: W. W. Norton.

Pissarides, Christopher A. 1988. "The Search Equilibrium Approach to Fluctuations in Employment." *American Economic Review* 78, no. 2: 363–68.

Rognlie, Matthew. 2015. "Deciphering the Fall and Rise in the Net Capital Share: Accumulation or Scarcity?" *Brookings Papers on Economic Activity*, Spring, 1–54.

Summers, Lawrence H. 1986. "Why Is the Unemployment Rate So Very High near Full Employment?" *Brookings Papers on Economic Activity*, no. 2: 339–83.

ROBERT J. GORDON
Northwestern University

Productivity and Growth over the Years at BPEA

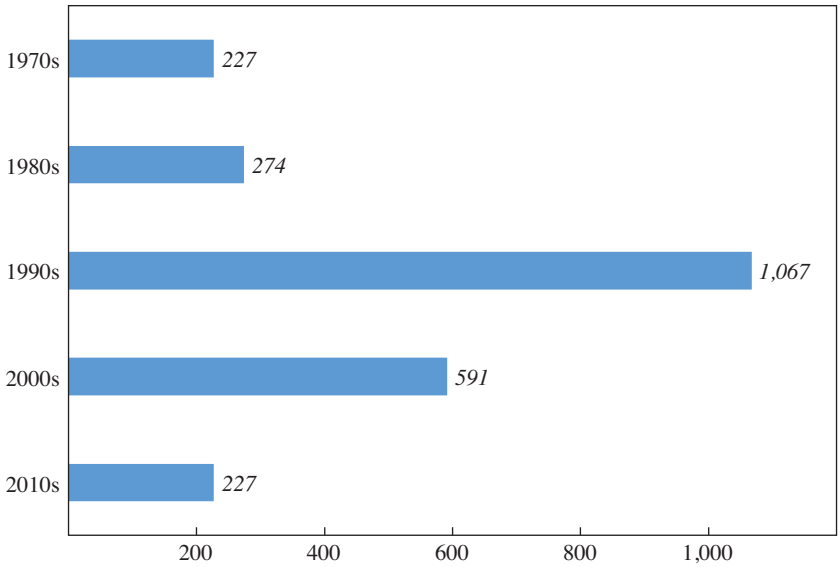
ABSTRACT Over the past fifty years *BPEA* authors have written extensively about both domestic US productivity growth and international sources of growth differences in rich and poor countries. This paper summarizes and evaluates five *BPEA* papers on US productivity growth that focus primarily on the sources of the post-1965 growth slowdown and post-1995 growth revival. Then three papers are reviewed on international growth differences, highlighting the difficulties of empirically determining the sources of growth and the competing roles in growth outcomes of structural factors like geography and demography versus policy and governmental issues, including legal systems, property rights, and absence of corruption.

Why did US productivity growth decelerate in the late 1960s, and why did it revive in the late 1990s? Why are some countries so rich and others so poor, and why do growth rates differ so much among nations? *BPEA* has a long history of concern with productivity as a source of growth in potential GDP for the US economy and also with economic growth more generally as it differentiates the world's more and less successful economies.

To choose among the many papers on these two topics—productivity growth in the United States and growth differences among nations—I have divided papers into two corresponding groups. The first concerns productivity and potential output growth in the US context. In this group I begin with two papers from the 1970s, written early in *BPEA*'s formative decade, and then follow with three papers written after 1980 that are chosen for

Conflict of Interest Disclosure: The author did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this paper. He is currently not an officer, director, or board member of any organization with an interest in this paper.

Figure 1. Mean Citations of *BPEA* Papers by Decade



Source: Author's calculations.

their relatively large number of citations. The second group includes papers on economic growth more generally as it differs among nations. Since there were no such papers in *BPEA*'s first decade I limit coverage to three papers written after 1980, also selected by the criterion of citations.

I. Citations and the History of *BPEA*

Because I used citation counts to choose among papers written after 1980, I couldn't help but notice a few interesting aspects of the citations. First is the inequality across decades, as shown by figure 1, a bar chart providing mean citations per paper for each of the five decades. It is very striking that the 1990s were the golden decade for *BPEA* citations, with the average paper receiving more than 1,000 citations. The 2000s come next, with the other decades far behind.

While I don't have a good explanation for the relatively low citation counts for the recent articles of the 2010s, other than their youth, I can suggest some aspects of the first decade of the 1970s that limited citations per

article. First was the short-run orientation of the early papers. Quoting from the editors' introduction to the first issue, "particular attention is devoted to recent and current economic developments that are directly relevant to the contemporary scene" (Okun and Perry 1970, 1).

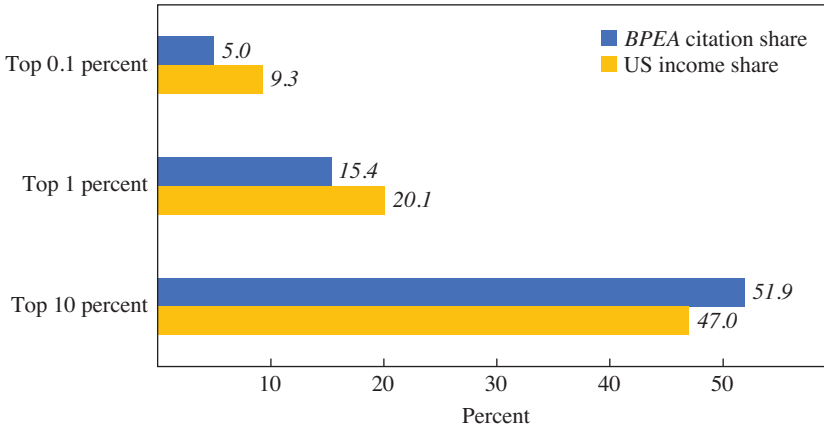
Another aspect was the initial conception of a panel of experts on particular topics that could be thought of as equations in a large-scale US macro-econometric model. Thus in the first two issues there were papers on components of aggregate demand—consumption, inventory investment, home building, and the federal budget—on monetary and fiscal policy, and on interactions between demand and supply in the form of papers on inflation and unemployment. This equation-by-equation approach tended to exclude a host of topics that did not fit into that framework.

Third, and perhaps most important in limiting citations per article, was that in the first year three of the papers in each issue were full length and the remaining four were short so-called sector reports, providing updates on topics for which authors had already written longer papers or would write longer papers in subsequent issues. Those sector reports were often quite short and were not assigned formal discussants. If sector reports were excluded, citation counts per paper for the 1970s would be considerably higher.

Not only are citation counts per article unequal by decade, but they are highly unequal across papers. The top 3 percent of papers, twenty out of the total of 717, accounted for about 20 percent of the total citations (the cutoff to make the top twenty is 1,239 citations; mean citations per article is 263, and median citations is 116).

This made me wonder, Is the distribution of *BPEA* citations more or less unequal than the distribution of US income? Taking data for 2014 from a paper by Piketty, Saez, and Zucman (2018), we obtain figure 2. Shown in three groups of bars are the percentage of citations and income accounted for by the top 0.1 percent, top 1 percent, and top 10 percent. In the first two groups income is slightly more unequally distributed than citations, but for the third group (10 percent) *BPEA* citations are a bit more unequally distributed than US income. Overall, we conclude that if *BPEA* authors were paid in proportion to their citations, the resulting inequality of *BPEA* authors' income would approximate the inequality of the US income distribution. We might also conclude that inequality of outcomes is inherent in many aspects of productive activity, from the stratospheric heights of wealth owned by the founders of today's internet giants to the more plebeian precincts of the Brookings Institution.

Figure 2. Distribution of *BPEA* Citations Compared to the Distribution of Income in the United States



Sources: Piketty, Saez, and Zucman (2018); author's calculations.

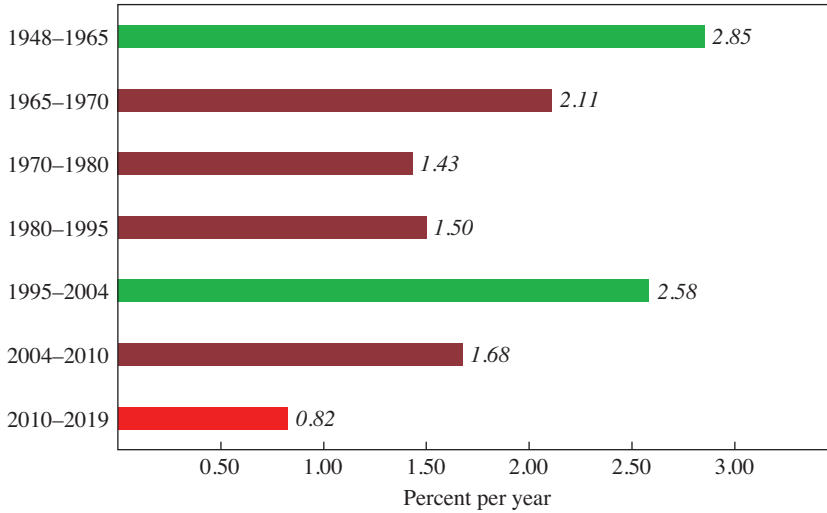
II. Productivity Growth

When *BPEA* began in early 1970 the United States had enjoyed relatively rapid growth in labor productivity and in potential GDP for many decades, and this was expected to continue. As I showed in my recent book, 1970 was the year that marked the end of a remarkable fifty years in which labor productivity growth in the total US economy had averaged 2.8 percent per year (Gordon 2016, 14). And productivity growth in the private sector was somewhat faster than that, roughly 3.2 percent, a number that had been codified as the acceptable rate of real wage growth in the program of wage-price guideposts of the Kennedy-Johnson era.[1] Thus *BPEA* began in an environment in which productivity growth around 3 percent was normal and could be expected to continue, and any recent shortfalls in observed productivity growth below 3 percent were worthy of note. As for potential GDP, as late as 1972 the official measure was estimated to be growing at 4.3 percent per year.[2]

1. Private business sector productivity growth was 3.2 percent between 1948 and 1965; see Baily and Gordon (1988, table 5). The current estimate from the BLS website is 3.17 percent for 1948–1965.

2. Nordhaus (1972, 526) cites the contemporaneous *Business Conditions Digest* as the source of the official estimate of potential GDP growth of 4.3 percent per annum.

Figure 3. Annual Growth Rate of Output per Hour, Total US Economy, 1948–2019, Selected Intervals



Source: Author's calculations.

Note: Output is measured as the geometric average of GDP and gross national income. Hours are an unpublished series provided by the Bureau of Labor Statistics.

For perspective both on the two productivity papers written in the early 1970s and those that later attempted to explain the post-1995 growth revival, figure 3 presents annual growth rates of US output per hour for selected intervals since 1948. It is important to note that these numbers refer to the total US economy, including the agriculture, government, and household sectors, not the more narrowly defined nonfarm private business (NFPB) sector that is the universe covered by the regularly published Bureau of Labor Statistics (BLS) quarterly productivity data.[3]

The attention here to the total economy rather than the NFPB sector reflects the coverage of the two early *BPEA* productivity papers reviewed here, both of which covered the total economy.

In figure 3 the pre-1970 period is divided at 1965, reflecting the break point chosen to mark the beginning of slower productivity growth that attracted the early *BPEA* papers on productivity growth, and shows that

3. Output per hour in the total economy is defined here as the average of gross domestic product (GDP) and gross domestic income (GDI) divided by hours of work in the total economy, an unpublished series that I obtain each quarter from the BLS.

growth slowed from 2.85 percent per annum in 1948–1965 to 2.11 percent in 1965–1970. Then came a long interval between 1970 and 1995 of even slower growth, broken into 1.43 percent for *BPEA*'s first decade and 1.50 percent for 1980–1995. Then arrived the remarkable revival to 2.58 percent during 1995–2004, followed by a two-step slowdown to 1.68 percent in 2004–2010 and the lamentable 0.82 percent rate recorded in the last prepandemic decade of 2010–2019.

The initial slowdown in 1965–1970 was soon noticed in the first *BPEA* paper to review the determinants of productivity and potential output growth, "Labor Force Structure, Potential Output, and Productivity," written by *BPEA* coeditor George Perry in 1971. He began by decomposing growth of the three main components of actual output growth—employment, hours per employee, and output per hour—over the 1948–1970 period divided up into three subintervals with breaks at 1955 and 1965. He showed that productivity growth had declined by more than half, from 3.4 percent in the first interval (1948–1955) to 1.6 percent in the third interval (1965–1970). Output growth had declined, but by less than productivity due to faster growth of employment. Notice that the 1.6 percent growth rate for 1965–1970 in Perry's contemporaneous data is substantially lower than the 2.1 percent growth rate for the same interval in today's retrospective data.

In order to understand the slowing trend, Perry highlighted the shift in the composition of the labor force toward two groups, women and teenagers, who were paid lower wages than adult men and worked fewer hours per week. He assumed that their lower observed wages reflected true differences in productivity and created new series that weighted each age-sex group by its relative wage and number of weekly hours, so women and teenagers were given a smaller weight. Then he took the age-sex weighted series for employment and hours per employee and created the cyclical adjustment needed to translate actual growth in employment, hours of work, and productivity into potential (i.e., cyclically adjusted) growth rates.

Perry's most striking finding was that *all* of the decline in the potential productivity growth of –0.4 percent per year could be attributed to the effect of the changing age-sex mix, so cyclically adjusted productivity with a constant age-sex mix would have grown at a constant rate of 2.9 percent per year over 1948–1970 with no slowdown. Since *actual* productivity growth as noted above had declined by 1.8 percent per year between the initial interval (1948–1955) and third interval (1965–1970), Perry's detailed analysis interpreted 0.4 percent of the 1.8 point slowdown as caused by the

age-sex effect, 0.7 to the cyclical impact of the large GDP gap of the terminal recession year 1970, and the remaining 0.7 percent as a residual error due to the unexplained low value of productivity in the year 1970.[4]

Because he explained away the large observed drop in productivity growth as the result of shifting age-sex weights and a cyclical effect in the terminal year 1970, and because he projected little further change in the age-sex mix, Perry concluded by extrapolating constant 2.9 percent productivity growth and 4.3 percent potential output growth into the future decade of the 1970s. We know in retrospect that his forecasts were too optimistic. Actual output growth between 1971 and 1980 was not his projected 4.3 percent per year but a much slower 3.2 percent; and, as shown in figure 3, actual productivity grew not at his projected 2.9 percent but at only half that rate, 1.43 percent.[5]

Eighteen months after Perry's paper, William Nordhaus tackled the slowing productivity growth conundrum in his 1972 paper, "The Recent Productivity Slowdown." This was entirely devoted to productivity growth without considering data on employment or hours per employee as had Perry. Nordhaus divided up the 1948–1971 period into three eras with the same 1955 and 1965 dividing points that Perry had chosen. In Nordhaus's data productivity growth slowed between the first interval (1948–1955) and the last interval (1965–1971) from 3.1 to 1.9 percent per year, an overall slowdown of 1.2 points, less than Perry's 1.8 point slowdown.[6]

In a brief survey of explanatory hypotheses for the observed productivity growth slowdown, Nordhaus considered but rejected Perry's hypothesis based on the age-sex employment mix. He disagreed with Perry's assumption that the lower wages of these groups reflected lower true productivity and instead argued that their lower wages reflected discrimination against them. Instead, he proposed and tested an industry composition hypothesis, that changes in the employment share of individual industries explained the slowdown, and he devised a decomposition that predicted a slowdown of 0.9 points out of the 1.2 points actually observed. His analysis isolated differences among industries in the level rather than the growth rate of

4. Perry (1971, 559) provides a decomposition for the single year 1970, and I have translated this into the implications for the slowdown in growth rates between the initial interval and the terminal interval.

5. The actual growth rate of 3.2 percent refers to the average of the current Bureau of Economic Analysis (BEA) estimates of GDP and GDI between 1970:Q4 and 1980:Q4.

6. Using current data, half of the difference in the measured slowdowns can be attributed to the 1965–1970 interval used by Perry and the 1965–1971 interval used by Nordhaus.

productivity as responsible for the slowdown and highlighted the role of agriculture and finance, insurance, and real estate (FIRE), and to a lesser extent durable manufacturing and the government sector. Can Perry's emphasis on the age-sex distribution be reconciled with Nordhaus's analysis of industry composition? Robert Solow, a formal discussant of both papers, suggested that much of the influx of women and teenagers was into low-productivity industries that contributed to Nordhaus's composition effect.

Like Perry, Nordhaus predicted future productivity growth through 1980. His forecast for 1972–1980 of 2.1 percent was the same as his cyclically corrected rate of 2.1 percent for 1965–1971. How accurate was that forecast? Today's data for 1972–1980 register a productivity growth rate of only 1.2 percent per year, so Nordhaus in retrospect was also too optimistic, although not as much as Perry.[7] From my perspective this is because both Perry's demographic hypothesis and Nordhaus's industrial composition hypothesis ignored the role of early postwar catch-up in the exploitation of what I have called the Great Inventions, the implementation of which had been delayed by the Great Depression and war. By this interpretation productivity growth slowed after the first postwar decade as this backlog of previous inventions worked its way through the production process.

We now turn to three papers on productivity that were written after 1980 and are chosen on the basis of citation counts. The first of these in chronological order was coauthored by Martin Baily and myself in 1988, "The Productivity Slowdown, Measurement Issues, and the Explosion of Computer Power." Written sixteen years after Nordhaus's contribution, the paper began by pointing to a productivity growth slowdown in the NFPB sector of 1.6 percentage points when 1973–1987 was compared to 1948–1973.[8]

Much of the analysis searched for measurement errors that could explain the slowdown. The claim was not that BEA and BLS had changed their methods to make measurement worse after 1973, but rather that the economy had changed in ways that made preexisting measurement errors more important. We emphasized that many measurement errors concern intermediate goods and just shuffle measured productivity growth among

7. Like Perry, Nordhaus's concept of productivity was for the total economy, including agriculture and government, not the NFPB sector. The 1.2 percent figure quoted in the text is the growth rate for 1971:Q4 to 1980:Q4 of the same series used to create figure 3.

8. The slowdown between 1948–1973 and 1973–1987 is 1.66 percent in today's published BLS data.

industries without explaining the aggregate slowdown. To make a contribution an error must influence the measurement of final goods output or total labor input.

We concluded that measurement issues could explain about one-third of the 1.6 point slowdown. This estimate of a 0.5 point measurement contribution combined errors in business services, airline fare discounts, and issues involving labor quality, including the age-sex adjustments that Perry emphasized together with evidence of declining labor quality based on test scores. We found plenty of other measurement errors, for example, for construction price deflators and unmeasured quality improvements in medical care, but these applied both before and after 1973 so did not help explain the slowdown. Much of our analysis unearthed measurement errors at the level of individual industries, and we emphasized that quality and convenience improvements in finance, communications, and transportation should more appropriately be credited to durable manufacturing. Overall, we concluded that the pattern of industry slowdowns was consistent with the "impetus to productivity advance in the early postwar years, perhaps a backlog of innovations and investment opportunities delayed by depression and war, followed, after the mid-1960s, by a depletion of opportunities" (Baily and Gordon 1988, 420). This is an early statement of the theme of innovation depletion that I have developed further more recently (Gordon 2000, 2016).

The last two papers on productivity growth, chosen by the citation criterion, were both published in 2002. The first of these marked William Nordhaus's return to the productivity topic in his paper "Productivity Growth and the New Economy," which shared with his paper of thirty years earlier an attention to the industry composition of productivity growth. To achieve this he constructed for 1977–2000 a new income-side database on output, hours worked, and productivity for each industry that added up to income-side total output or GDI. This allowed him to distinguish between a pure productivity effect that sums the industries with fixed output shares, a Baumol effect of shifting output shares, and a Denison effect of the interaction between shifting shares of hours and output. In contrast to his 1972 paper, which attributed most of the post-1965 productivity growth slowdown to shifting shares, the new paper found the Baumol effect to be near zero throughout the post-1977 interval.

Nordhaus's most important and widely cited result was that, in contrast to most research on the post-1995 productivity revival by Jorgenson, Oliner, Sichel, and others, the acceleration was not entirely or even primarily driven by the new economy information and communications

technology (ICT) sector, which includes computer hardware, communications, and software.[9] Instead, Nordhaus showed that the revival in 1995–2000 as compared to 1977–1995 had an ICT contribution of 33 percent for his preferred income-side measure, of only 17 percent for the conventional nonfarm business sector product-side concept, and 38 percent for the subset of industries that he classified as "well-measured."[10] The headline result of a 17 percent new economy contribution thus leaves as unexplained the majority of the post-1995 productivity growth revival.

In the usual *BPEA* fashion, the discussant remarks shed substantial light on the sources of the differences between Nordhaus's relatively small ICT contribution to the revival and the much larger contribution attributed by other authors. Leaving aside technicalities including income-side versus product-side concepts, time period definitions, and data revisions, the most important source of the reconciliation was the limitation by Nordhaus of the ICT contribution to the ICT-producing industries without counting at all the contribution of ICT capital to productivity growth in the ICT-using industries.

Speaking of ICT, the final productivity paper reviewed here also appeared in the 2002 *BPEA*: "Intangible Assets: Computers and Organizational Capital" by Erik Brynjolfsson, Lorin Hitt, and Shinkyu Yang. Written in the midst of the 1995–2004 revival in productivity growth highlighted in figure 3 that many had attributed to an upsurge of investment in ICT capital, the authors investigated the relationship at the firm level between computers and their payoff in the form of faster productivity growth and higher firm market value. Their basic message was that the effectiveness of computers depended on changes in firm organization and business practices. Their study, based on data for hundreds of firms over eleven years and limited to computer-using firms rather than those creating computer hardware or software, interpreted organizational assets as being much like other types of assets that contribute to long-term growth in output, productivity, profits, and market value.

One of their most striking findings was that financial markets placed substantially more value on installed computer capital than on other types of capital. The extent of that additional valuation depended on the implementation of reorganization—use of teams and team-based incentives, more

9. For example, see Jorgenson and Stiroh (2000) and Oliner and Sichel (2000).

10. These percentages come from my discussion of Nordhaus (2002, 248). The larger contribution of the new economy for the well-measured portion of the economy occurs because all of the new economy subsectors are considered to be well-measured.

broadly defined jobs, individual decision-making authority, and investment in skills and education. Firms with higher levels of both computer investment and these organizational characteristics had a higher market value and higher measured productivity than firms that invested only in computers or only in organizational change.

A big issue that concerned the authors and discussants was the huge size of the regression coefficient on computer capital—$1 in computer capital produced $15 in market valuation—and the addition of the organization variable did not decrease this by much. This anomaly led to consideration of reverse causation in the form of firms with highly successful business models, including internal reorganization, having plenty of market value available to buy computer capital. The paper and its discussants cited the example of Wal-Mart as a highly valued and productive company which combined organizational change, including the big-box store format, with wide-ranging investment in computers that allowed much greater control of inventories and the supply chain than had previously been possible.

III. The Growth of Nations

We now turn to three highly cited papers written between 1995 and 2003 that approach the topic of economic growth more generally as an inquiry into the reasons for differences in growth rates across countries and the failure of poor countries to converge to the output per capita level of rich countries. The continuing gigantic gap in living standards between rich and poor countries has been called the most important topic in economics, and Robert Lucas (1988) once famously wrote that when one starts thinking about it, "it is hard to think about anything else" (5).

The first in this group of three is Gregory Mankiw's 1995 paper, "The Growth of Nations." Mankiw contrasted the enormous differences in standards of living between a group of rich countries including Germany, Japan, and the United States and a group of poor countries including India, Indonesia, and Nigeria. What were the fundamental factors that made some countries so rich and others so poor? Mankiw noted that this question had long been neglected, but by the time of his paper in 1995, it had emerged as a subject in economics as important as the study of business cycles.

Mankiw's paper began by enumerating the central well-known deficiencies of the neoclassical Solow growth model when contrasted with reality—it predicted lower magnitudes of international income differences, faster rates of conditional convergence, and larger differences in the marginal product of capital. However, if capital's share in the Cobb-Douglas

formulation was around two-thirds instead of one-third, these problems of matching theory with reality faded away, and a consideration of human capital as well as externalities from physical capital easily justified a high capital share. Regarding endogenous growth theory, Mankiw made the bold claim that it added limited value to cross-country studies, because knowledge models were hard to check with international data. In addition, endogenous growth models did not apply well to East Asia, where the primary source of growth was capital accumulation rather than total factor productivity (TFP).

Turning to cross-country regressions, Mankiw noted several difficulties. As models became more subtle it became harder to distinguish among them empirically. For instance multiple models predicted conditional convergence. More important, cross-country regressions suffered from simultaneity (explanatory variables were the result of growth themselves), multicollinearity (explanatory variables were too closely correlated among themselves), and low degrees of freedom (too many variables and too few years of observation).

Mankiw concluded by finding the neoclassical Solow model still useful when supplemented by a broader view of capital that raised the numerical value of capital's share and when joined by endogenous growth theory as a useful supplement to identify the sources of increases in knowledge. But he ended on a pessimistic note. The neoclassical emphasis on differences in capital accumulation as the source of growth outcomes shifted the spotlight to the question of why some countries saved and invested so much more than others, and little progress had been made to answer that question.

Mankiw was just as pessimistic about policy implications. If capital accumulation was the key to growth, then policymakers should encourage more saving and investment from domestic and foreign sources. But beyond that economists had not developed persuasive methods of measuring the externalities from capital accumulation, and the lack of such measurements could lead to "haphazard policy, which is surely worse than no policy at all." Further, "policymakers who want to foster economic growth would do well to heed the first rule for physicians: do no harm" (Mankiw 1995, 309).

The second highly cited paper to tackle differing growth rates across nations appeared in *BPEA* in 1998—"Geography, Demography, and Economic Growth in Africa" by David Bloom and Jeffrey Sachs. The authors took a different approach from many studies of economic growth in Africa that focused on macroeconomic policy, market liberalization, and institutions. Instead they created a convincing case for geography and

demography as factors that had substantially limited economic growth in Africa.

Geography was a hindrance because the climate near the equator was humid, temperatures were high, and there was no monsoon to provide irrigation. These factors severely limited agricultural productivity, leading African countries to specialize in cash crops (coffee, mangoes) that were suitable for the climate, requiring much food to be imported. The hot and humid climate was a natural host for infectious diseases, such as malaria and yellow fever, which took a direct economic toll and deterred foreign settlement and investment.

Geography did not just involve climate but also topography. Africa lacked deep harbors as dot the coastlines of Europe and North America and also in some regions lacked great navigable rivers. Thus transportation costs were high, made worse by the fact that most of Africa's population lived inland where numerous countries were landlocked. Both the presence of malaria and being landlocked were isolating, and isolation was a major cause of slow growth.

As if geography were not a sufficient barrier to growth, Africa also suffered from unfavorable demography. A combination of high fertility rates with better public health practices that had improved survival rates had led to rapid population growth and a high ratio of dependent youths. A larger population strained the availability of natural resources, while the youth of the population limited saving and investment. Solutions to excess population growth were difficult to achieve, as contraceptives were not widely available. Further, Africans actively desired large families in part due to persistent social norms and a lack of education.

Bloom and Sachs concluded that policy and governance were not the most important factors impeding the achievement of more rapid economic growth in Africa. They argued that causality ran strongly from geography, demography, and public health to growth with little reverse causation. Their rough estimate was that two-thirds of the explanation of Africa's slow growth could be traced to these underlying structural factors and only the remaining one-third to economic policy and institutions. They lamented the relative lack of international research on tropical health issues and the relationship between geography and agricultural productivity.

Bloom and Sachs noted that Africa was the only region in the world to experience an absolute decline in real exports per capita between 1980 and 1996. They called for a major shift, particularly in coastal cities, to the types of low capital intensive manufactured exports that had formed the backbone of rapid growth in East Asia. They also emphasized the need

for infrastructure, which could be financed privately rather than by "cash-strapped state monopolies" (Bloom and Sachs 1998, 272). In good *BPEA* fashion, the discussants strongly disagreed. One pointed to a whole array of issues that were amenable to policy changes, including dictatorship, civil wars, a lack of electricity, poor contract enforcement, poor information caused partly by an abysmal telephone system, and the perception of foreign investors of a high political risk of expropriation.

The last of the trilogy of highly cited growth papers is the 2003 contribution by Barry Bosworth and Susan Collins, "The Empirics of Growth: An Update." The authors examined two methods of studying growth—growth accounting and regressions—and attempted to reconcile widely divergent findings from these two methods regarding the sources of growth across nations. Findings differed on whether physical and capital accumulation were the main underlying sources of growth or whether the main source was advances in TFP.

Bosworth and Collins pointed to measurement issues that largely explained these different conclusions. For instance it mattered whether the capital stock was directly evaluated or whether it was approximated from investment rates, a practice of which the authors disapproved. Likewise the failure to find an association between educational attainment and output growth may have reflected measurement errors in educational quality. The authors examined several data sets on educational attainment used in previous studies and found them poorly correlated with each other.

The main contribution of the paper was to improve the measurement of the key variables by constructing a new data set for eighty-four countries accounting for 95 percent of world GDP over the four-decade interval of 1960–2000. The authors concluded that both growth accounting and regressions were useful tools, conditional on correct measurement of variables, cleaning up differences in data and definitions, and—in regression analyses—inclusion of previously omitted variables. They found that the contribution of capital had been understated in some previous studies due to the flawed practice of using investment series as a proxy for capital input. Somewhat surprisingly, they found a limited role for educational attainment, partly due to the difficulty of finding adequate measures of educational quality.

Besides an emphasis on capital accumulation, Bosworth and Collins pointed to initial conditions and government policy as important explanations of high or low growth. There was a strong negative correlation between growth and initial per capita income, supporting conditional convergence. Life expectancy in the initial year as a measure of health also had

a significant positive correlation with growth. Governmental institutions were strongly correlated with growth, including law and order, absence of corruption, and protection of property rights. While this emphasis on institutions went against the structural handicap hypothesis of Bloom and Sachs, those authors were supported by the finding that a tropical climate hindered growth.

Bosworth and Collins supplemented their positive conclusions with several negatives—there appeared to be no relationship between growth and either macroeconomic policies or openness to trade. And they admitted that their set of variables shed little light on one of the main puzzles treated in their paper, the sharp slowdown in world growth from the two decades before 1980 to the two decades after 1980 in most of the world outside of India and China.

IV. Conclusion

As shown in figure 3, growth in labor productivity in the United States is delineated by four postwar eras—fast 1948–1965, slow 1965–1995, fast 1995–2004, and slow again after 2004 with a second wave of retardation after 2010. Some combination of a changing age-sex mix and altered industry composition, as in the early Perry (1971) and Nordhaus (1972) papers, makes a partial contribution to understanding the initial phase of the slowdown from pre-1955 to 1965–1970. But the overoptimistic forecasts of these authors for the decade of the 1970s, which were based on holding constant the age-sex and industrial shares, suggest that something more profound was going on. My more recent suggestion (Gordon 2000, 2016) that this omitted factor was diminishing returns to innovation and depletion of important innovations developed in a fruitful earlier era, was originally set forth in the paper by Baily and Gordon (1988) summarized above.

The 1995–2004 productivity growth revival is widely attributed to the invention of small powerful computers and of the internet, together with an explosion of investment in ICT capital (at least through the year 2000). Nordhaus in 2002 understated the contribution of ICT capital but pointed to something important, the strong post-1995 revival of industries that were not intensive users of ICT equipment. His position became stronger in the data that emerged after he wrote, from 2002 to 2004, when productivity growth remained strong despite a sharp decline of ICT investment.

As for the fourth era of slowdown after 2004, and particularly since 2010, we return to the diminishing returns argument, this time applied to

the wave of ICT innovation that reached its peak in the late 1990s. But at least two qualifications need to be introduced to the implied pessimism about future US productivity growth. First, there is very little correlation between productivity growth in one decade versus the next, and so we could be on the cusp of another revival propelled by robots and artificial intelligence. Second, Brynjolfsson, Collis, and Eggers (2019) have recently shown that the invention of the smartphone and tablet have produced vast amounts of consumer surplus that can be measured in consumer surveys, raising questions about the adequacy of conventional output measures.

The papers on international growth range from Mankiw's (1995) skepticism that empirical research can uncover the sources of growth, to Bloom and Sachs's (1998) insistence that structural impediments are more important than policy or institutions, to the demonstration by Bosworth and Collins (2003) that regressions can provide a convincing decomposition of the sources of growth. As a spectator to the international growth literature, I emerge from this review with some frustration that even the best efforts of Bosworth and Collins could not explain why worldwide growth slowed down from 1980 to 2000, not to mention why India and China were exceptions to that slowdown. The sources of the stunning growth achievement of East Asia over more than four successive decades still raise questions about the relative role of investment, government oversight, and culture. For future generations of *BPEA* authors there are plenty of new puzzles that arise in the experience since 2000, including why worldwide growth in emerging economies regained momentum, and why it surged even in a significant number of African countries.

ACKNOWLEDGMENTS I am grateful to Ziqiao Zhang for his help in preparing this paper.

References

Baily, Martin Neil, and Robert J. Gordon. 1988. "The Productivity Slowdown, Measurement Issues, and the Explosion of Computer Power." *Brookings Papers on Economic Activity*, no. 2: 347–431. https//doi.org/10.2307/2534534.

Bloom, David E., and Jeffrey D. Sachs. 1998. "Geography, Demography, and Economic Growth in Africa." *Brookings Papers on Economic Activity*, no. 2: 207–95. https://doi.org/10.2307/2534695.

Bosworth, Barry P., and Susan M. Collins. 2003. "The Empirics of Growth: An Update." *Brookings Papers on Economic Activity*, no. 2: 113–79. www.jstor.org/stable/1209192.

Brynjolfsson, Erik, Avinash Collis, and Felix Eggers. 2019. "Using Massive Online Choice Experiments to Measure Changes in Well-Being." *Proceedings of the National Academy of Sciences* 116, no. 15: 7250–55. https://doi.org/10.1073/pnas.1815663116.

Brynjolfsson, Erik, Lorin M. Hitt, and Shinkyu Yang. 2002. "Intangible Assets: Computers and Organizational Capital." *Brookings Papers on Economic Activity*, no. 1: 137–98. www.jstor.org/stable/1209176.

Gordon, Robert J. 2000. "Does the New Economy Measure Up to the Great Inventions of the Past?" *Journal of Economic Perspectives* 14, no. 4 (Fall): 49–74.

Gordon, Robert J. 2016. *The Rise and Fall of American Growth.* Princeton, N.J.: Princeton University Press.

Jorgenson, Dale W., and Kevin J. Stiroh. 2000. "Raising the Speed Limit: U.S. Economic Growth in the Information Age." *Brookings Papers on Economic Activity*, no. 1: 125–211.

Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22:3–42.

Mankiw, N. Gregory. 1995. "The Growth of Nations." *Brookings Papers on Economic Activity*, no. 1: 275–326. https://doi.org/10.2307/2534576.

Nordhaus, William D. 1972. "The Recent Productivity Slowdown." *Brookings Papers on Economic Activity*, no. 3: 493–545. https://doi.org/10.2307/2534126.

Nordhaus, William D. 2002. "Productivity Growth and the New Economy." *Brookings Papers on Economic Activity*, no. 2: 211–65. www.jstor.org/stable/1209207.

Okun, Arthur M., and George L. Perry. 1970. "Editors' Introduction." *Brookings Papers on Economic Activity*, no. 1: 1–7.

Oliner, Stephen D., and Daniel E. Sichel. 2000. "The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?" *Journal of Economic Perspectives* 14, no. 4: 3–22.

Perry, George L. 1971. "Labor Force Structure, Potential Output, and Productivity." *Brookings Papers on Economic Activity*, no. 3: 533–78. https://doi.org/10.2307/2534156.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. "Distributional National Accounts: Methods and Estimates for the United States." *Quarterly Journal of Economics* 133, no. 2: 553–609.

ALAN S. BLINDER
Princeton University

BPEA *and Monetary Policy over Fifty Years*

ABSTRACT Ever since its first issue in 1970, *BPEA* has played a leading role in the analysis of monetary policy. This paper surveys *BPEA*'s many contributions to three specific areas: (1) the Phillips curve, which provides the empirical bridge between real economic activity and inflation; (2) the analysis and demise of monetarism, the doctrine that emphasized the money supply over interest rates; and (3) evaluations of and recommendations for actual monetary policy in the United States, which began in the first *BPEA* issue and continues to this day. *BPEA* has played a dominant (though not monopoly) role in each of these areas.

In thinking about the historic role the Brookings Panel has played as an intermediary and incubator of ideas between the academic world and the world of actual monetary policy, it is critical to remember both the intellectual and policy settings when *BPEA* began in 1970.

On the intellectual front, academic macroeconomics was far less theoretical and far more grounded in reality than it has been in recent decades. Giant econometric models, built rather loosely on a Keynesian theory that was itself loose, roamed the earth. In fact, one such dinosaur inhabited the Brookings Institution. Theoretical looseness was tolerated in those days.

While the Keynesian paradigm dominated the policy world, the monetarist-Keynesian wars were raging—both in academia and in some central banking circles. A lively debate on the subject between Milton Friedman and Walter Heller (1969) had taken place at New York University in November 1968, the same year that Karl Brunner (1968) coined the term "monetarism." The then famous paper by Andersen and Jordan (1968), which purported to show empirically that money growth mattered for GDP

Conflict of Interest Disclosure: The author did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this paper. He is currently not an officer, director, or board member of any organization with an interest in this paper.

but fiscal variables did not, had sparked controversy and consternation both in the academy and outside it. Although the intellectual world didn't know it yet, it was awaiting William Poole's (1970b) seminal paper on money supply targeting versus interest rate targeting, which was sitting in the publication queue at the *Quarterly Journal of Economics*.

Perhaps most important, the subsequent view that macro stabilization policy *is* monetary policy, period, had not yet taken hold. Fiscal policy was thought of as at least a coequal partner, and perhaps even as the senior partner. In the policy world, both monetary policy and fiscal policy had turned contractionary to fight inflation in the late 1960s—the former joining the fight eagerly, the latter reluctantly. I believe the income tax surcharge of 1968—studied by Arthur Okun (1971) in one of the earliest Brookings papers—was the first and last time contractionary fiscal policy was deliberately used to slow the US economy.[1] The 1969–1970 recession which followed was in progress when the first Brookings Panel convened. Then, as since, the conveners of the panel did not much like recessions.

Okun, who had chaired President Johnson's Council of Economic Advisers until January 1969, and George Perry, recently arrived from the University of Minnesota, made a fantastic team. Together, they recruited an all-star cast for the inaugural Brookings Panel. Its members included some who were older but skewed decidedly young—featuring Poole (b. 1937), William Branson (b. 1938), Robert J. Gordon (b. 1940), Barry Bosworth (b. 1942), and Robert Hall (b. 1943). Okun and Perry had an eye for talent (and in case you're wondering, I'm younger than all those guys). That first Brookings Panel also included, as senior advisers, such luminaries as Lawrence Klein, Paul Samuelson, and Robert Solow—not to mention a business consultant named Alan Greenspan. Taken in toto, this list evokes John F. Kennedy's quip about a 1962 gathering of Nobel Prize winners at the White House being the greatest collection of brainpower to dine there since Thomas Jefferson dined alone.

Perry, of course, is still part of the Brookings Panel, and we tip our hats to him today.

Early meetings of the panel basically covered the Keynesian waterfront as mapped out in the macro textbooks of the day. There were papers on consumption, investment, the government budget, money demand, and net exports—and, of course, on the Phillips curve. Almost all of that was relevant to monetary policy, but I will confine myself here to three

1. There were subsequent fiscal contractions, but they were motivated by bringing down the budget deficit, not by slowing down the economy.

prominent topics: the Phillips curve, money growth and monetarism, and ideas for and evaluations of monetary policy.

I. The Phillips Curve

Monetary policy is in large measure about the control of inflation, including the linkages between the real side of the economy (e.g., output and employment) and the nominal side (e.g., money and inflation). So it was altogether fitting and proper that the first paper at the first meeting of the Brookings Panel was on the Phillips curve, which links the two. It was written by Gordon (1970), who was just thirty at the time, and turned out to be the first in a long series of papers by Gordon on the Phillips curve in *BPEA*. Indeed, the names Gordon, Brookings, and Phillips will be linked forever in the history of macroeconomic thought.

Once again, it is important to remember the intellectual setting in 1970. A. W. Phillips's (1958) original paper had used wage inflation as the left-hand variable and basically dismissed inflation, not to mention expected inflation, as a right-hand variable. This omission was not an oversight. Phillips (1958, 283) argued that "cost of living adjustments will have little or no effect on the rate of change of money wage rates." Really? Two years later, Phillips's colleague Richard Lipsey (1960) remedied that deficiency by estimating an inflation coefficient of 0.37 in a wage Phillips curve of the form:

$$w_t = \alpha\pi_t + f(U_t) + \varepsilon_t,$$

where w_t is the rate of change of nominal wages, $f(U_t)$ is a nonlinear function of the unemployment rate, π_t is the inflation rate, and ε_t is a stochastic error term. When Lipsey (1960) estimated that same equation with more modern data, rather than Phillips's 1861–1913 sample, his estimate of α rose to 0.76 (with standard error 0.08). Much higher, but still significantly below 1.

The view in 1970 was that, while Friedman's (1968) and Phelps's (1967, 1968) theoretical arguments for why α should be 1 were persuasive, the data showed $\alpha < 1$.[2] For example, that first *BPEA* paper by Gordon (1970) estimated α to be just 0.45.[3] It was as Groucho Marx might have put

2. Thomas Sargent's (1971) brilliant little paper, which showed why $\alpha = 1$ was beside the point under rational expectations, was not yet appreciated.

3. For this equation, Gordon (1970, 36–37) used an auxiliary equation for nominal bond rates to estimate expected inflation as a function of past inflation rates.

it if he had a PhD in economics: "Who ya gonna believe, Milton or your lyin' eyes?"

Soon, however, empirical eyesight improved, largely through Gordon's efforts in *BPEA*. By the second 1972 meeting, he already had an estimated Phillips curve with a nonlinear α coefficient that rose as expected inflation rose, reaching 1 at an expected inflation rate around 7 percent (Gordon 1972). Thus, by 1972 or 1973, the empirical/theoretical conflict over the verticality of the long-run Phillips curve was all but over.[4] It was vertical both in theory and in practice.

But the Gordon-*BPEA*-Phillips curve saga was far from over. The first big postwar supply shocks hit in 1972–1973, driving inflation far above what Phillips curves without supply shocks predicted.[5] As CPI inflation in the United States rose from 3.4 percent in 1972 (December to December) to 8.9 percent in 1973 and 12.1 percent in 1974—during a recession, no less!—monetarists crowed that Keynesian economics, with its misguided Phillips curve, was inherently inflationary. A few years later, Lucas and Sargent (1978, 49) chimed in that the "predictions" of Keynesian economics "were wildly incorrect, and that the doctrine on which they were based is fundamentally flawed" so "the task which faces contemporary students of the business cycle is that of sorting through the wreckage." Wow! And that was just on the first page.

The Brookings Panel was not persuaded, however; it kept the Keynesian embers glowing. The main inflationary villain at the meetings was not profligate Keynesian government spending, but rather supply shocks. Months before the Organization of the Petroleum Exporting Countries (OPEC) struck, Bosworth and Farmer (1973) called attention to crop failures, disappearing anchovies, and the resulting food price explosion as proximate sources of inflation. A year later, Popkin (1974, 259) concluded that "the effect of commodity inflation was substantial in 1973." More fundamentally, in that same issue, Pierce and Enzler (1974) of the Federal Reserve Board staff modified the Keynesian MIT-Penn-SSRC (MPS) model to analyze the macroeconomic impacts of what they called "external inflationary shocks." Their simulations showed stagflation, of course: output fell and inflation rose.[6]

4. The debate over whether the short-run Phillips curve was vertical was still several years away.

5. For a full discussion of those early supply shocks, see Blinder (1982) or Blinder and Rudd (2013).

6. Well, not quite. Their main simulations held *nominal* money supply growth constant, meaning that *real* money growth fell, which eventually extinguished the inflation.

Notice, please, that all this analysis came very quickly—far faster than the scholarly journals could react. Speedy publication has always been an important advantage of *BPEA*; the *Journal of Political Economy* never specialized in current events.

By the first issue of 1975, Gordon (1975a) was back with a clear conceptual analysis of supply shocks that was quite similar to what Phelps (1978) would publish three years later. And two issues after that, Gordon (1975b) presented his first Phillips curve that fully incorporated supply shocks. I remember well that Nordhaus (1975, 663), in discussing that paper, referred to it as "Chateau Gordon 1975." It was a good vintage, though not Gordon's last.

Out of this early work—and including also contributions by Nordhaus (1972), Perry (1970), Schultze (1971), and others—came what I have long called the Brookings Rule of Thumb—that each point-year of unemployment above the natural rate reduced inflation by half a percentage point. It's a rule that worked well in the United States for decades. In the mid-1990s, as vice chairman of the Federal Reserve, I routinely referred to the fine performance of the Phillips curve as "the clean little secret of macro econometrics."[7] As a matter of fact, I still use the Brookings Rule of Thumb to show my Economics 101 students that contemporary estimates of the Phillips curve give an almost perfect explanation of the Volcker disinflation. You don't need any magical credibility effects or M2 growth rates.

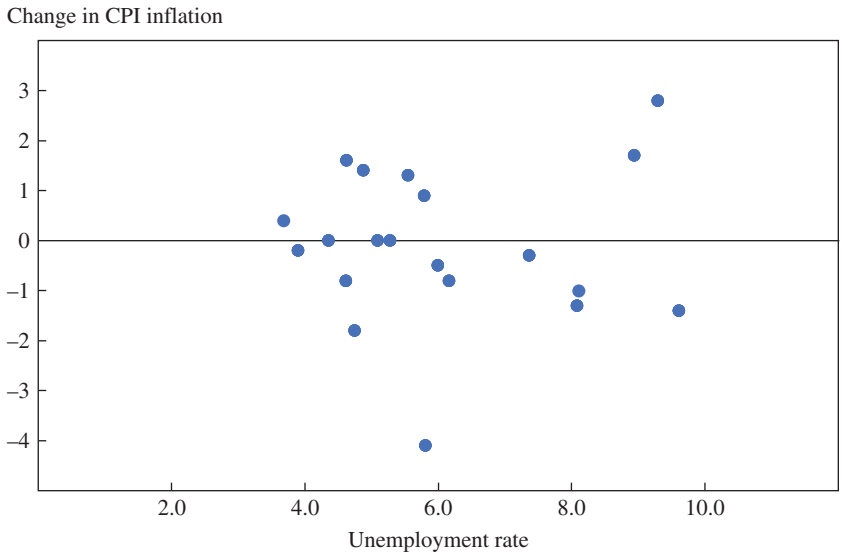
Phillips curve research went relatively quiet in *BPEA* after the early 1980s, with just two papers that concentrated on Phillips curves in the late 1980s, one by Blanchard (1987) and one by Ball, Mankiw, and Romer (1988).[8] A fascinating paper by Akerlof, Dickens, and Perry (1996)—which the trio followed up with in Akerlof, Dickens, and Perry (2000)—shook the intellectual tree a bit by adding what I'd call "non-Gordonesque" aspects, such as extreme downward wage rigidity and money illusion.

Starting with Chateau Gordon 1998, *BPEA* papers began grappling with the empirical failure of the Phillips curve. The first question was: Why didn't the low unemployment rates of the late 1990s raise inflation more? Gordon (1998) partly patched things up by incorporating several new supply shocks and by adapting the idea of a time-varying non-accelerating inflation rate of unemployment (NAIRU) from Staiger, Stock, and Watson

7. See, for example, my "notorious" (to some) Jackson Hole speech (Blinder 1994, 340).

8. For this count, and in what follows, I interpret the phrase "concentrating on Phillips curves" fairly strictly. It excludes, for example, many related papers on labor market developments, which I leave to Hall's paper in this issue.

Figure 1. Change in CPI Inflation versus Unemployment Rate, 2001–2020



Source: US Bureau of Labor Statistics.

(1997).[9] Katz and Krueger (1999) subsequently estimated the effects on NAIRU of several labor market developments—such as demographic change and mass incarceration.

The second question arose after the Great Recession: Why didn't such a deep recession reduce inflation more? Krueger, Cramer, and Cho (2014), echoing Gordon (2013), argued that part of the explanation was that the long-term unemployed exert much less downward pressure on wage inflation than the short-term unemployed. Ball and Mazumder (2011) suggested that the slope of the Phillips curve varied over time. But by the time you've allowed both the intercept (the NAIRU) and the slope to change over time, you haven't got much of a Phillips curve left. And we didn't.

Notice that both of these questions suggest a flatter Phillips curve—as does the scatter plot in figure 1. Suffice it to say that the Brookings Rule of Thumb no longer works, and the stable Phillips curve is no longer a "clean little secret." Its failure is well known.

9. On new supply shocks, see Blinder and Yellen (2001).

II. Money Growth and Monetarism

Monetarism, which became popular in the 1960s and early 1970s, combined the positive doctrine that changes in the growth rate of money dominated changes in the growth rate of nominal (and perhaps even real) GDP with the normative doctrine that central banks would do their monetary policy duties better if they kept the money supply growing at a constant (and modest) rate. In the early days of *BPEA*, Okun and Perry seemed to take stamping out the scourge of monetarism as part of their mission. And they succeeded—with much help from both academic papers, some of which appeared in *BPEA* (see below), and real-world events. I still remember that when Steve Goldfeld and I were working on what became my first published paper (Goldfeld and Blinder 1972), either Okun or Perry insisted that we include what became a one-page "Digression on the Behavior of the Money Supply." Yes, it was off topic. But to those two missionaries, it was right on point.

I examined the historical record to see how many *BPEA* papers focused on money growth and monetarism. In doing so, I applied a strict filter, excluding papers that were mainly about interest rates, exchange rates, bank regulation, or the savings and loan debacle, even though all of these bear on monetary policy. To get into my count, a paper had to focus on the relationship between money growth and GDP growth, the instability of money demand, or the role of financial innovation therein. There were a whopping twenty-five such papers in the 1970s alone, and six more in the 1980s. Of these, Goldfeld's (1973, 1976) two papers on money demand stand out. Lest you think the *BPEA* editors didn't brook dissent, six of those twenty-five papers were authored or coauthored by Poole, the house monetarist.

The mention of Poole leads straight to two historical ironies.

First, although Poole was a monetarist himself, it was his seminal 1970 paper that laid the intellectual groundwork for the eventual demise of monetarism. Poole (1970b) used an extremely simple—and therefore intuitively transparent—model to show that money supply targeting is preferred when IS shocks dominate macro fluctuations, but interest rate targeting is preferred when LM shocks dominate. As time went by in the 1970s and 1980s, it became abundantly clear that LM shocks were gigantic, in the United States and elsewhere, presumably because financial innovations kept funds sloshing around from one definition of M to another.

Although Poole obtained his central finding in an extremely simple fixed-price model, it proved to be remarkably robust. In fact, although the

connection seems to have been all but forgotten, Poole's paper led directly to the famous Sargent and Wallace (1975) paper, which held even more dire implications for academic views on monetary policy. Given the importance of Sargent and Wallace (1975) in the history of macroeconomic thought in the academy, it is worth remembering that the central point of their paper was that adding rational inflationary expectations to Poole's model carried stunning implications. We all know where Sargent and Wallace (1975) led. But this is not the place to review the long, acrimonious debates over new classical economics because they took place mostly outside of *BPEA*.

The second big irony is that high inflation, the root of monetarism's ascendancy during the 1960s and 1970s, wound up accounting for its demise in the 1970s. As inflation rose in the late 1960s, Friedman and other monetarists successfully branded Keynesianism as inherently inflationary. That was effective public relations, but the charge wasn't true. In fact, both Heller and Okun, as CEA chairs, had urged President Johnson to raise taxes as a way to first head off, and later to reduce, demand-pull inflation from Vietnam spending. But Johnson didn't want anything to interfere with his grand plan to prosecute the war in Vietnam and the war on poverty at the same time. As always, politics triumphed over economics in the policy arena. But in the intellectual market, Keynesian stock sunk and monetarist stock rose.

Later, Lucas and Sargent (1978, 51) upped the ante, declaring Keynesian models to be guilty of "econometric failure on a grand scale" for much the same reason: inflation rose. This time, while there was a small dose of demand-pull inflation in, say, 1977–1978, the main culprits were a series of food and energy shocks that the rational expectations school somehow ignored. (Was doing that really rational?)

In October 1979, Federal Reserve Chair Paul Volcker announced the Federal Reserve's putative conversion to monetarism. Was it genuine? I'm pretty sure Volcker was not an avid reader of either Lucas and Sargent (1978) or *BPEA*. His wonderful memoir (Volcker 2018, 118) makes it clear that his conversion to monetarism was mainly a mechanism for tying the Federal Reserve Open Markets Committee (FOMC) to the inflation-fighting mast—and also a better way to explain the fight to the general public.

Years before Volcker's chairmanship, the high inflation of the late 1960s and 1970s had interacted badly with nominal interest rate ceilings and other corsets on banks, thereby incentivizing wave after wave of financial innovation designed to elide dysfunctional regulations. Seeing such LM shocks happening on a grand scale, one central bank after another abandoned either the pretense or practice of monetarism. (At the Federal

Reserve, it seemed mostly to be pretense.) As Gerald Bouey, the governor of the Bank of Canada at the time, famously quipped, "We did not abandon M1, M1 abandoned us."[10] So where monetarism was concerned, it was: inflation giveth, and inflation taketh away.

Appropriately, the Brookings Panel turned its attention to financial innovation and the instability of money demand early and often. Goldfeld's two papers in the 1970s were already mentioned; the second (Goldfeld 1976) was provocatively titled, "The Case of the Missing Money."[11] Among other things, that paper discussed financial innovations as causes of the decline in money demand. Two *BPEA* issues earlier, the Federal Reserve's Enzler, Johnson, and Paulus (1976, 279) had "speculated that much of the weakness in money demand reflects innovations and regulatory changes." It was sound speculation even though Poole, in discussing their paper, was unconvinced. The next year, with Poole again the discussant, Friedman (1977) wrote provocatively (to monetarists) about "The Inefficiency of Short-Run Monetary Targets for Monetary Policy." Them's fightin' words.

Attention turned to financial innovation in earnest at the first Brookings Panel meeting of 1979, when another team from the Federal Reserve (Porter, Simpson, and Mauskopf 1979) presented a paper titled, "Financial Innovation and the Monetary Aggregates." Their analysis held little good news for using the monetary aggregates, although Poole (1979), in an accompanying paper, was still unconvinced. And don't forget that 1979 was the momentous year the Federal Reserve turned putatively to monetarism.

The final *BPEA* nails in the monetarist coffin were hammered in by Hester (1981), Lindsey (1982), and Simpson (1984) in the early 1980s. Hester (1981, 142) emphasized that "monetary policy is poorly designed if it fails to take into account the possibility that conditions which result from policy changes may lead to innovations." I was the discussant of Simpson's paper, and my opening words (Blinder 1984, 266) summarized it as "an intelligent brief about why the Federal Reserve should not have done what it did between October 1979 and October 1982." By the time Bosworth (1989) penned his paper titled "Institutional Change and the Efficacy of Monetary Policy" and Romer and Romer (1990) wrote "New Evidence on the Monetary Transmission Mechanism," monetarism was not even mentioned. Okun was probably smiling from the grave. Perry was probably smiling in his seat.

10. Canada, House of Commons Standing Committee on Finance, Trade and Economic Affairs, Minutes of Proceedings and Evidence, no. 134, March 28, 1983, 12.

11. Full disclosure: I believe I suggested that title.

III. Advice for Monetary Policymakers

Analysis and evaluation of monetary policy in *BPEA* did not, of course, end with debunking monetarism. It has been a focus of the Brookings Panel from its earliest days to today. In examining this voluminous literature, I applied another strict filter, restricting myself to papers that clearly either evaluated monetary policy decisions (generally, the Federal Reserve's) or dispensed advice to monetary policymakers. This filter excludes, for example, many interesting and important papers on financial crises (not just the big one), bank regulation, credit controls, and the like. It also excludes a number of notable "big think" papers that are highly relevant to monetary policy, such as Okun's (1973) famous "Upward Mobility in a High-Pressure Economy," Blanchard and Simon's (2001) early paper on the Great Moderation, and Sims's (2002) insightful "The Role of Models and Probabilities in the Monetary Policy Process." My filter nonetheless left a whopping fifteen Brookings papers in the 1970s, five in the 1980s, seven in the 1990s, nine in the 2000s, and fifteen in the 2010s. Since that adds up to fifty-one, I'll just hit some highlights.

The Brookings Panel has never shied away from giving advice to monetary policymakers. That tradition started in the first issue of *BPEA* with a short paper by Kareken (1970, 161), who concluded by observing that "the implication would seem to be that the economy may take one course if the FOMC uses the [Treasury] bill rate and money market variables in specifying policy, as it did in 1969, and another if it uses one or more of the monetary aggregates." I wonder if Okun and Perry put him up to that.

Jump all the way to the fall 2018 issue, and you'll find two papers offering advice to the Federal Reserve. One was written by a team from the Boston Federal Reserve that included its president, Eric Rosengren (Fuhrer and others 2018), and set the stage for the Federal Reserve's subsequent review of its strategy, tools, and communications. The other was a symposium on policy at the effective lower bound, which featured a contribution from Yellen (2018), in which she advocated a lower-for-longer strategy for short rates similar to what the Federal Reserve had promulgated during her chairmanship.[12] So here was a case of a former Federal Reserve chair using *BPEA* to give advice to a current chair.

But back to history. In the second *BPEA* issue, Poole (1970a, 273) examined, and seemed to laud, gradualism in fighting inflation. In his words,

12. The issue also included short papers by Forbes (2018), Hamilton (2018), and Swanson (2018).

"politicians and the informed public now clearly recognize that excessive zeal in fighting inflation will produce excessive unemployment." Right! But this is not a message I associate with either monetarists or new classical economists. I do, however, associate it with empirical Phillips curves featuring sticky prices.

Furthermore, Poole (1970c) was back in the following issue with a paper titled "Whither Money Demand?" which examined the econometric difficulties of estimating a demand-for-money function. Was Poole shunning his role as the house monetarist? No. He soon bolstered his monetarist credentials with a long paper on how the Federal Reserve could and should improve its control of the money stock (Poole and Lieberman 1972). Perusing those early *BPEA* volumes, it is hard to escape the conclusion that Poole was overworked.

In 1972, Okun (1972) provided a thorough and thoughtful examination of what we now call the rules versus discretion debate. That paper came years before Kydland and Prescott (1977), but long after Friedman (1948). Friedman and the monetarists, of course, were arguing for a k-percent rule for money growth, basing their case largely on imperfect knowledge of the economy and imperfect behavior by policymakers. Okun (1972, 157) concluded at the time that "rules for fixed instrument settings would not achieve our objectives. . . . The proponents of rules . . . have provided good questions and bad answers." Much the same could be said today, except that today's rules don't have "fixed instrument settings."

The k-percent rule fell of its own weight when monetarism collapsed. It was replaced by Kydland and Prescott's (1977) argument that central bankers have an inflation bias—itself a dubious proposition—and that tying their hands with rules is the way to correct it. Their argument was further developed by Barro and Gordon (1983) and others, and it had enormous influence within academia—but not, I believe, in central banks. The popularity of these time inconsistency models in academia was somewhat amazing, given what was happening in the real world at the time. The models basically predicted that inflation would always be too high, not that it would rise (as it had from 1965 to 1980 in the United States) and then fall (as it did after 1980). Did time inconsistency somehow get worse and then get better?

The third incarnation of the rules versus discretion debate, which is still with us today, revolves around the Taylor (1993) rule. It was taken up in Kocherlakota's (2016) fascinating paper—of which I was a discussant (Blinder 2016). Kocherlakota's conclusions were (a) that it seems unlikely on basic theoretical grounds that an inevitably imperfect rule would be

superior to inevitably imperfect discretion, and (b) that the Taylor rule, in particular, may have led the FOMC to be too timid in pushing the economy out of the Great Recession.

Returning to the 1970s, the first *BPEA* issue of 1974 featured a debate between Tobin (1974), perhaps the leading Keynesian of the day, and Poole (1974) over what the Federal Reserve should do to end the deep recession. You probably can guess what each gentleman said. But you probably can't guess the names of the two discussants: the father-and-son team of Robert Aaron Gordon and Robert J. Gordon (1974). Almost poetic.

The following year was notable for the paper by Modigliani and Papademos (1975) that coined the term non-inflationary rate of unemployment (NIRU, later corrected to NAIRU) and offered estimates thereof ranging from 5.1 percent to 5.8 percent. They advised the Federal Reserve that, as the economy struggled its way out of the deep recession, "monetary policy should be aimed at explicitly stated targets for real output and employment" (Modigliani and Papademos 1975, 141). Nominal anchors were not yet in vogue.

The previously discussed preoccupation with monetarism dominated the 1970s and 1980s. So I'll skip ahead to 1990, when Romer and Romer (1990) published a sequel to the "narrative approach" they had pioneered in Romer and Romer (1989). Their focus at the Brookings Panel meeting that day was comparing the conventional IS-LM view of how monetary policy works (via bank reserves and money) with the so-called lending or credit view, which emphasizes the unique importance of bank loans. Their reading of the evidence favored the former, but that was thirty years ago.

A year later, Bernanke, who was destined for greater things, teamed up with Lown of the New York Federal Reserve to write a widely cited paper on the credit crunch of 1990 (Bernanke and Lown 1991). It would not be Bernanke's last notable Brookings paper. In 1997, he partnered with Gertler and Watson to write what some people view as the definitive analysis of oil shocks and monetary policy (Bernanke, Gertler, and Watson 1997). In 2004, while a governor of the Federal Reserve, Bernanke, Reinhart, and Sack (2004) presented an important assessment, "Monetary Policy Alternatives at the Zero Bound," that is frequently cited on this issue which is still very much alive. Several years after he retired from the Federal Reserve, Bernanke (2018) was in a better position than almost anyone else to assess the real effects of disrupted credit during the financial crisis. The panel audience was all ears that day. Interestingly, but not surprisingly, he placed great emphasis on the credit view that the Romers had debunked in 1990. *BPEA* is not monolithic.

But back to history—way baaack. In 1998, Krugman (1998) created a stir, and subsequently a boatload of citations, with his famous paper, "It's Baaack: Japan's Slump and the Return of the Liquidity Trap." That paper was the first of what would become a series of Brookings papers dealing in one way or another with the "zero" lower bound on nominal interest rates. Five years later, Eggertsson and Woodford (2003) published their famous paper on optimal monetary policy at the zero lower bound, which made the case for price-level targeting. That paper, along with Reifschneider and Williams (2000), is often credited with being the inspiration for the "lower for longer" idea that the Federal Reserve adopted in 2013.

Williams (2009) was the research director at the San Francisco Federal Reserve when he addressed the zero lower bound question at the fall 2009 Brookings Panel meeting. He suggested that day that the 2 percent inflation target might be too low—a conclusion that, given his current position, he may want to blame on his identical twin. The zero lower bound issue was also addressed, in a wide variety of ways, by Edge and Gürkaynak (2010), Swanson (2011), Krishnamurthy and Vissing-Jorgensen (2011), Svensson (2011), Campbell and others (2012), Chodorow-Reich (2014), Evans and others (2015), Kiley and Roberts (2017), and, as mentioned earlier, a fall 2018 symposium featuring Yellen, Forbes, Hamilton, and Swanson. Whew! Some of these papers focused on forward guidance or quantitative easing. Notice that many of the authors on this list were either staff members or decision makers of the Federal Reserve.

It is no exaggeration to say that *BPEA* has been one of the main outlets for research and writing on unconventional monetary policy. Brookings was also exploiting its comparative advantage on speed here; the more academic journals were much slower.

IV. After Fifty Years

So, as we look back today on fifty years of writing about and debate over monetary policy in *BPEA*, what are the major contributions of the panel?

Most clearly, I think, the Brookings Panel has played a dominant—though not a monopoly—role in the development and evolution of empirical Phillips curves. Gordon was clearly the leader in this domain, though he had plenty of help; and I look forward to sampling Chateau Gordon 2022 once he's figured it all out.

On the demise of monetarism, which was one of the presumed original goals of Okun and Perry, you might say the job was easy: monetarism fell of its own weight. But it didn't always look that way in real time, and

Goldfeld's (1976) missing money plus a host of *BPEA* papers on financial innovations and money demand played significant roles.

When it came to thinking sensibly about supply shocks, inflation, and monetary policy, I'd say that *BPEA* was there early and often while too many academic economists were not—and indeed are still not. And on monetary policy more generally, I'd emphasize, as *BPEA* standouts, the defense of discretion against rules, the great attention given to estimates of the NAIRU, and the spate of ideas on how to cope with (or to avoid) the zero lower bound.

More fundamentally, I'd argue, the Brookings Panel kept Keynesian ideas alive and kicking through onslaughts first from monetarism, then from new classical economics, real business cyclists, and even supply-side economics. Over the decades, *BPEA* has been consistently less faddish, and more closely tied to the earth, than the major academic journals. Today's Keynesianism differs in many ways from Keynes's *General Theory* (1936), and also from what you can read in the early issues of *BPEA*—as it should. But it remains the best game in town.

References

Akerlof, George A., William T. Dickens, and George L. Perry. 1996. "The Macroeconomics of Low Inflation." *Brookings Papers on Economic Activity*, no. 1: 1–76.

Akerlof, George A., William T. Dickens, and George L. Perry. 2000. "Near-Rational Wage and Price Setting and the Long-Run Phillips Curve." *Brookings Papers on Economic Activity*, no. 1:1–60.

Andersen, Leonall C., and Jerry L. Jordan. 1968. "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization." *Federal Reserve Bank of St. Louis Review* 50: 11–24.

Ball, Laurence, N. Gregory Mankiw, and David Romer. 1988. "The New Keynesian Economics and the Output-Inflation Trade-Off." *Brookings Papers on Economic Activity*, no. 1: 1–82.

Ball, Laurence, and Sandeep Mazumder. 2011. "Inflation Dynamics and the Great Recession." *Brookings Papers on Economic Activity*, Spring, 337–405.

Barro, Robert J., and David B. Gordon. 1983. "A Positive Theory of Monetary Policy in a Natural Rate Model." *Journal of Political Economy* 91, no. 4: 589–610.

Bernanke, Ben S. 2018. "The Real Effects of Disrupted Credit: Evidence from the Global Financial Crisis." *Brookings Papers on Economic Activity*, Fall, 251–322.

Bernanke, Ben S., Mark Gertler, and Mark Watson. 1997. "Systematic Monetary Policy and the Effects of Oil Price Shocks." *Brookings Papers on Economic Activity*, no. 1: 91–157.

Bernanke, Ben S., and Cara S. Lown. 1991. "The Credit Crunch." *Brookings Papers on Economic Activity*, no. 2: 205–47.

Bernanke, Ben S., Vincent R. Reinhart, and Brian P. Sack. 2004. "Monetary Policy Alternatives at the Zero Bound: An Empirical Assessment." *Brookings Papers on Economic Activity*, no. 2: 1–78.

Blanchard, Olivier Jean. 1987. "Aggregate and Individual Price Adjustment." *Brookings Papers on Economic Activity*, no. 1: 57–122.

Blanchard, Olivier Jean, and John Simon. 2001. "The Long and Large Decline in U.S. Output Volatility." *Brookings Papers on Economic Activity*, no. 1: 135–64.

Blinder, Alan S. 1982. "The Anatomy of Double Digit Inflation in the 1970s." In *Inflation: Causes and Effect*, edited by Robert E. Hall. Chicago: University of Chicago Press.

Blinder, Alan S. 1984. Comment on "Changes in the Financial System: Implications for Monetary Policy," by Thomas S. Simpson. *Brookings Papers on Economic Activity*, no. 1: 266–72.

Blinder, Alan S. 1994. "Overview." In *Economic Policy Symposium Proceedings: Reducing Unemployment: Current Issues and Policy Options*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Blinder, Alan S. 2016. Comment on "Rules versus Discretion: A Reconsideration," by Narayana Kocherlakota. *Brookings Papers on Economic Activity*, Fall, 37–41.

Blinder, Alan S., and Jeremy B. Rudd. 2013. "The Supply-Shock Explanation of the Great Stagflation Revisited." In *The Great Inflation: The Rebirth of Modern Central Banking*, edited by Michael D. Bordo and Athanasios Orphanides. Chicago: University of Chicago Press.

Blinder, Alan S., and Janet L. Yellen. 2001. *The Fabulous Decade: Macroeconomic Lessons from the 1990s*. New York: Century Foundation Press.

Bosworth, Barry P. 1989. "Institutional Change and the Efficacy of Monetary Policy." *Brookings Papers on Economic Activity*, no. 1: 77–124.

Bosworth, Barry P., and John Farmer. 1973. "The Current Inflation: Malign Neglect?" *Brookings Papers on Economic Activity*, no. 1: 263–83.

Brunner, Karl. 1968. "The Role of Money and Monetary Policy." *Federal Reserve Bank of St. Louis Review*, July, 4–22.

Campbell, Jeffrey R., Charles L. Evans, Jonas D. M. Fisher, and Alejandro Justiniano. 2012. "Macroeconomic Effects of Federal Reserve Forward Guidance." *Brookings Papers on Economic Activity*, Spring, 1–80.

Chodorow-Reich, Gabriel. 2014. "Effects of Unconventional Monetary Policy on Financial Institutions." *Brookings Papers on Economic Activity*, Spring, 155–204.

Edge, Rochelle M., and Refet S. Gürkaynak. 2010. "How Useful Are Estimated DSGE Model Forecasts for Central Bankers?" *Brookings Papers on Economic Activity*, Fall, 209–59.

Eggertsson, Gauti B., and Michael Woodford. 2003. "The Zero Bound on Interest Rates and Optimal Monetary Policy." *Brookings Papers on Economic Activity*, no. 1: 139–211.

Enzler, Jared, Lewis Johnson, and John Paulus. 1976. "Some Problems of Money Demand." *Brookings Papers on Economic Activity*, no. 1: 261–80.

Evans, Charles, Jonas Fisher, François Gourio, and Spencer Krane. 2015. "Risk Management for Monetary Policy near the Zero Lower Bound." *Brookings Papers on Economic Activity*, Spring, 141–96.

Forbes, Kristin. 2018. "Monetary Policy at the Effective Lower Bound: Less Potent? More International? More Sticky?" *Brookings Papers on Economic Activity*, Fall, 521–42.

Friedman, Benjamin M. 1977. "The Inefficiency of Short-Run Monetary Targets for Monetary Policy." *Brookings Papers on Economic Activity*, no. 2: 293–346.

Friedman, Milton. 1948. "A Monetary and Fiscal Framework for Economic Stability." *American Economic Review* 38, no. 3: 245–64.

Friedman, Milton. 1968. "The Role of Monetary Policy." *American Economic Review* 58, no. 1: 1–17.

Friedman, Milton, and Walter W. Heller. 1969. *Monetary vs Fiscal Policy*. New York: W. W. Norton.

Fuhrer, Jeff, Giovanni P. Olivei, Eric S. Rosengren, and Geoffrey M. B. Tootell. 2018. "Should the Federal Reserve Regularly Evaluate Its Monetary Policy Framework?" *Brookings Papers on Economic Activity*, Fall, 443–97.

Goldfeld, Stephen M. 1973. "The Demand for Money Revisited." *Brookings Papers on Economic Activity*, no. 3: 577–646.

Goldfeld, Stephen M. 1976. "The Case of the Missing Money." *Brookings Papers on Economic Activity*, no. 3: 683–739.

Goldfeld, Stephen M., and Alan S. Blinder. 1972. "Some Implications of Endogenous Stabilization Policy." *Brookings Papers on Economic Activity*, no. 3: 585–644.

Gordon, Robert J. 1970. "The Recent Acceleration of Inflation and Its Lessons for the Future." *Brookings Papers on Economic Activity*, no. 1: 8–47.

Gordon, Robert J. 1972. "Wage-Price Controls and the Shifting Phillips Curve." *Brookings Papers on Economic Activity*, no. 2: 385–421.

Gordon, Robert J. 1975a. "Alternative Responses of Policy to External Supply Shocks." *Brookings Papers on Economic Activity*, no. 1: 183–206.

Gordon, Robert J. 1975b. "The Impact of Aggregate Demand on Prices." *Brookings Papers on Economic Activity*, no. 3: 613–70.

Gordon, Robert J. 1998. "Foundations of the Goldilocks Economy: Supply Shocks and the Time-Varying NAIRU." *Brookings Papers on Economic Activity*, no. 2: 297–346.

Gordon, Robert J. 2013. "The Phillips Curve Is Alive and Well: Inflation and the NAIRU during the Slow Recovery." Working Paper 19390. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w19390/w19390.pdf.

Gordon, Robert J., and Robert Aaron Gordon. 1974. "Discussion of Tobin and Poole Reports." *Brookings Papers on Economic Activity*, no. 1: 246–48.

Hamilton, James D. 2018. "The Efficacy of Large-Scale Asset Purchases When the Short-Term Interest Rate Is at Its Effective Lower Bound." *Brookings Papers on Economic Activity*, Fall, 543–54.

Hester, Donald D. 1981. "Innovations and Monetary Control." *Brookings Papers on Economic Activity*, no. 1: 141–99.

Kareken, John H. 1970. "Monetary Policy in 1969." *Brookings Papers on Economic Activity*, no. 1: 152–62.

Katz, Lawrence F., and Alan B. Krueger. 1999. "The High-Pressure U.S. Labor Market of the 1990s." *Brookings Papers on Economic Activity*, no. 1: 1–87.

Keynes, John Maynard. 1936. *General Theory of Employment, Interest and Money*. New York: Harcourt, Brace.

Kiley, Michael T., and John M. Roberts. 2017. "Monetary Policy in a Low Interest Rate World." *Brookings Papers on Economic Activity*, Spring, 317–72.

Kocherlakota, Narayana. 2016. "Rules versus Discretion: A Reconsideration." *Brookings Papers on Economic Activity*, Fall, 1–36.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." *Brookings Papers on Economic Activity*, Fall, 215–87.

Krueger, Alan B., Judd Cramer, and David Cho. 2014. "Are the Long-Term Unemployed on the Margins of the Labor Market?" *Brookings Papers on Economic Activity*, Spring, 229–80.

Krugman, Paul R. 1998. "It's Baaack: Japan's Slump and the Return of the Liquidity Trap." *Brookings Papers on Economic Activity*, no. 2: 137–205.

Kydland, Finn E., and Edward C. Prescott. 1977. "Rules Rather Than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy* 85, no. 3: 473–91.

Lindsey, David E. 1982. "Recent Monetary Developments and Controversies." *Brookings Papers on Economic Activity*, no. 1: 245–71.

Lipsey, Richard G. 1960. "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1862–1957: A Further Analysis." *Economica* 27, no. 105: 1–31.

Lucas Jr., Robert E., and Thomas J. Sargent. 1978. "After Keynesian Macroeconomics." In *After the Phillips Curve: Persistence of High Inflation and High Unemployment*. Conference Series no. 19. Boston: Federal Reserve Bank of Boston.

Modigliani, Franco, and Lucas Papademos. 1975. "Targets for Monetary Policy in the Coming Year." *Brookings Papers on Economic Activity*, no. 1: 141–65.

Nordhaus, William D. 1972. "The Worldwide Wage Explosion." *Brookings Papers on Economic Activity*, no. 2: 431–65.

Nordhaus, William D. 1975. Comment on "The Impact of Aggregate Demand on Prices," by Robert J. Gordon. *Brookings Papers on Economic Activity*, no. 3: 663–65.

Okun, Arthur M. 1971. "The Personal Tax Surcharge and Consumer Demand, 1968–70." *Brookings Papers on Economic Activity*, no. 1: 167–211.

Okun, Arthur M. 1972. "Fiscal-Monetary Activism: Some Analytical Issues." *Brookings Papers on Economic Activity*, no. 1: 123–72.

Okun, Arthur M. 1973. "Upward Mobility in a High-Pressure Economy." *Brookings Papers on Economic Activity*, no. 1: 207–61.

Perry, George L. 1970. "Changing Labor Markets and Inflation." *Brookings Papers on Economic Activity*, no. 3: 411–48.

Phelps, Edmund S. 1967. "Phillips Curves, Expectations of Inflation and Optimal Unemployment over Time." *Economica* 34, no. 135: 254–81.

Phelps, Edmund S. 1968. "Money-Wage Dynamics and Labor-Market Equilibrium." *Journal of Political Economy* 76, no. 4: 678–711.

Phelps, Edmund S. 1978. "Commodity-Supply Shock and Full-Employment Monetary Policy." *Journal of Money, Credit and Banking* 10, no. 2: 206–21.

Phillips, A. W. 1958. "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957." *Economica* 25, no. 100: 283–99.

Pierce, James L., and Jared J. Enzler. 1974. "The Effects of External Inflationary Shocks." *Brookings Papers on Economic Activity*, no. 1: 13–61.

Poole, William. 1970a. "Gradualism: A Mid-Course View." *Brookings Papers on Economic Activity*, no. 2: 271–301.

Poole, William. 1970b. "Optimal Choice of Monetary Policy Instruments in a Simple Stochastic Macro Model." *Quarterly Journal of Economics* 84, no. 2: 197–216.

Poole, William. 1970c. "Whither Money Demand?" *Brookings Papers on Economic Activity*, no. 3: 485–501.

Poole, William. 1974. "Reflections on U.S. Macroeconomic Policy." *Brookings Papers on Economic Activity*, no. 1: 233–46.

Poole, William. 1979. "The Monetary Deceleration: What Does It Mean and Why Is It Happening?" *Brookings Papers on Economic Activity*, no. 1: 231–38.

Poole, William, and Charles Lieberman. 1972. "Improving Monetary Control." *Brookings Papers on Economic Activity*, no. 2: 293–342.

Popkin, Joel. 1974. "Commodity Prices and the U.S. Price Level." *Brookings Papers on Economic Activity*, no. 1: 249–59.

Porter, Richard D., Thomas D. Simpson, and Eileen Mauskopf. 1979. "Financial Innovation and the Monetary Aggregates." *Brookings Papers on Economic Activity*, no. 1: 213–29.

Reifschneider, David, and John C. Williams. 2000. "Three Lessons for Monetary Policy in a Low-Inflation Era." *Journal of Money, Credit and Banking* 32, no. 4: 936–66.

Romer, Christina D., and David H. Romer. 1989. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." *NBER Macroeconomics Annual* 4:121–70.

Romer, Christina D., and David H. Romer. 1990. "New Evidence on the Monetary Transmission Mechanism." *Brookings Papers on Economic Activity*, no. 1: 149–213.

Sargent, Thomas J. 1971. "A Note on the Accelerationist Controversy." *Journal of Money, Credit and Banking* 3, no. 3: 721–25.

Sargent, Thomas J., and Neil Wallace. 1975. "'Rational' Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule." *Journal of Political Economy* 83, no. 2: 241–54.

Schultze, Charles L. 1971. "Has the Phillips Curve Shifted? Some Additional Evidence." *Brookings Papers on Economic Activity*, no. 2: 452–67.

Simpson, Thomas D. 1984. "Changes in the Financial System: Implications for Monetary Policy." *Brookings Papers on Economic Activity*, no. 1: 249–72.

Sims, Christopher A. 2002. "The Role of Models and Probabilities in the Monetary Policy Process." *Brookings Papers on Economic Activity*, no. 2: 1–40.

Staiger, Douglas, James H. Stock, and Mark W. Watson. 1997. "The NAIRU, Unemployment and Monetary Policy." *Journal of Economic Perspectives* 11, no. 1: 33–49.

Svensson, Lars E. O. 2011. "Practical Monetary Policy: Examples from Sweden and the United States." *Brookings Papers on Economic Activity*, Fall, 289–352.

Swanson, Eric T. 2011. "Let's Twist Again: A High-Frequency Event-Study Analysis of Operation Twist and Its Implications for QE2." *Brookings Papers on Economic Activity*, Spring, 151–207.

Swanson, Eric T. 2018. "The Federal Reserve Is Not Very Constrained by the Lower Bound on Nominal Interest Rates." *Brookings Papers on Economic Activity*, Fall, 555–72.

Taylor, John B. 1993. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy* 39: 195–214.

Tobin, James. 1974. "Monetary Policy in 1974 and Beyond." *Brookings Papers on Economic Activity*, no. 1: 219–32.

Volcker, Paul A. 2018. *Keeping at It: The Quest for Sound Money and Good Government*. New York: Hachette Book Group.

Williams, John C. 2009. "Heeding Daedalus: Optimal Inflation and the Zero Lower Bound." *Brookings Papers on Economic Activity*, Fall, 1–37.

Yellen, Janet. 2018. "Comments on Monetary Policy at the Effective Lower Bound." *Brookings Papers on Economic Activity*, Fall, 573–88.

WILL DOBBIE
Harvard University

CRYSTAL YANG
Harvard University

The Economic Costs of Pretrial Detention

ABSTRACT We measure the economic costs of the US pretrial system using several complementary approaches and data sources. The pretrial system operates as one of the earliest points of entry in the criminal justice system. It typically represents an individual's first opportunity to be incarcerated, potentially leading to subsequent long-term damage in the form of family separation, work interruption, loss of housing, and so on. We find that individuals lose almost $30,000 in forgone earnings and social benefits when detained in jail while awaiting the resolution of their criminal cases. These adverse consequences are also present in aggregate measures of economic well-being, with increases in county pretrial detention rates associated with increases in poverty rates and decreases in employment rates. Counties with high levels of pretrial detention also exhibit significantly lower levels of intergenerational mobility among children, consistent with pretrial detention having an adverse impact on young children who may be the dependents of individuals affected by the pretrial system.

The US criminal justice system has experienced a more than threefold expansion in the past several decades, with the number of inmates in local jails, state prisons, federal prisons, and privately operated facilities rising from 220 per 100,000 US residents in 1980 to 756 per 100,000 US residents in 2008. In addition, the high and growing incarceration rate in the United States has had a disproportionate impact on economically disadvantaged and minority communities, with significantly higher arrest, conviction, and incarceration probabilities for Black and Hispanic individuals compared to observably similar white individuals (Abrams, Bertrand, and

Conflict of Interest Disclosure: The authors did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this paper. They are currently not officers, directors, or board members of any organization with an interest in this paper.

Mullainathan 2012; Anwar, Bayer, and Hjalmarsson 2012; Rehavi and Starr 2014; McConnell and Rasul 2018; Raphael and Rozo 2019). By some estimates, more than 8 percent of the adult population and 33 percent of the Black adult male population in the United States has a prior felony conviction (Shannon and others 2017).

The economic consequences of mass incarceration may be substantial. Criminal records can result in substantial barriers to employment, particularly for minority individuals (Pager 2003; Holzer, Raphael, and Stoll 2006, 2007; Agan and Starr 2018), and recent work has shown that the increase in nonwork among US men, and Black men in particular, is partially attributable to the rising incarceration rate (Western 2002; Moffitt 2012; Neal and Rick 2014; Bayer and Charles 2018). These findings are particularly concerning given the persistently lower employment rates for Black and Hispanic individuals compared to non-Hispanic whites, with some of these gaps increasing in the wake of economic recessions and the recent COVID-19 pandemic.

In this paper, we measure the economic costs of the US pretrial system—an important but often neglected feature of the US criminal justice system that affects more than 10 million individuals each year who are arrested for an offense in the United States. To put this annual number into perspective, consider that approximately 8.1 million jobs were lost during the Great Recession between December 2007 and November 2009 (Shierholz 2009). The pretrial system operates as one of the earliest points of entry in the criminal justice system and typically represents an individual's first opportunity to be incarcerated. In theory, the pretrial system is meant to ensure the equitable release of most individuals before trial while minimizing the risk of flight or danger to the public. Despite this, defendants detained before trial represent over 75 percent of all jail inmates in some parts of the country. The pretrial system has also faced increased public scrutiny in recent years due to the all too common stories of arrested individuals who, despite being first-time offenders accused of low-level crimes, spend months in pretrial detention and face subsequent long-term damage in the form of family separation, work interruption, loss of housing, or even death. Pretrial detention can also generate substantial spillover effects, as the costs of paying money bail and other related court fees and fines often fall on other family and community members of detained individuals. Many of these harms can result even when the period of incarceration is brief and individuals are not ultimately convicted of any crime.

We measure the economic costs of the US pretrial system using several complementary approaches and data sources. We begin by describing

causal evidence about the effects of pretrial detention on individual economic outcomes such as formal labor market attachment and the receipt of unemployment insurance (UI) and the earned income tax credit (EITC), drawing on estimates from Dobbie, Goldin, and Yang (2018). Using quasi-experimental estimates based on the random assignment of cases to bail judges, these estimates capture the direct effects of detention on "marginal" defendants, or those for whom judges disagree on the appropriateness of pretrial detention. These quasi-experimental estimates show that individuals are nearly 10 percentage points less likely to be employed in the formal labor market if detained before trial. Detained individuals are also significantly less likely to receive EITC payments and receive substantially smaller UI and EITC amounts. Taken together, the estimates from Dobbie, Goldin, and Yang (2018) imply that individuals lose an average of $29,000 over the course of the working-age life cycle when detained in jail for just three days while awaiting the resolution of their criminal cases. These individual-level estimates further suggest that reforms such as the elimination of money bail could potentially increase aggregate earnings by as much as $80.91 billion per year, although we caution that the underlying quasi-experimental estimates are measured with considerable noise.

Our second contribution is to show that the adverse consequences of pretrial detention remain present in aggregate measures of economic well-being that also include potential spillover effects on other individuals in the household or community at large. One potential limitation of the quasi-experimental estimates from Dobbie, Goldin, and Yang (2018) is that they estimate direct effects on marginally detained defendants, whether through a short-run incapacitation or long-run job destabilizing effect. But pretrial detention is likely to generate spillover effects on other individuals in both the short and long run given its potential impact on families and communities. These spillover effects are all the more likely given that more than half of individuals detained pretrial are parents of children under the age of 18.[1] In the absence of convincing quasi-experimental variation in aggregate pretrial detention rates, we estimate the aggregate effects of pretrial detention inclusive of these spillover effects by comparing changes in county-level poverty and employment-to-population rates to changes in county-level pretrial detention rates. While the analysis is exploratory in nature, we find that a 10 percentage point increase in county pretrial

1. Based on data from the Survey of Inmates in Local Jails, at Prison Policy Initiative, "How Does Unaffordable Money Bail Affect Families?," https://www.prisonpolicy.org/blog/2018/08/15/pretrial/.

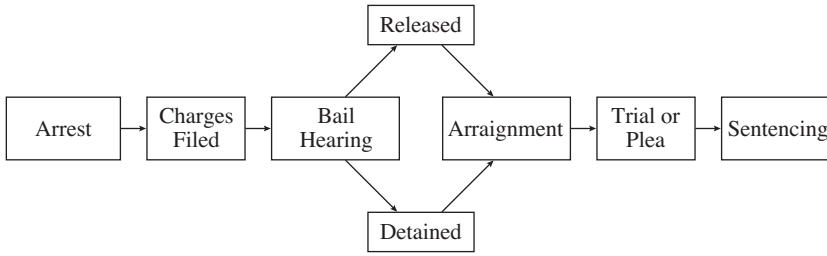
detention rates between 2000 and 2009 is associated with a 1.41 percentage point increase in county poverty rates and a 2.06 percentage point decrease in county employment rates between 2000 and 2010. The general relationship between changes in pretrial detention rates and poverty and employment rates is similar when we add additional controls and is typically stronger for economic outcomes of demographic groups most likely to be affected by pretrial detention, such as working-age Black individuals.

The final contribution of our paper is to explore more tentatively the intergenerational spillover effects of pretrial detention on young children. Leveraging measures of intergenerational mobility obtained from Chetty and others (2018), we find that counties with high levels of pretrial detention when children are young (age 7–12) exhibit significantly lower levels of intergenerational mobility for children when they are in adulthood. We find, for example, that a 10 percentage point higher pretrial detention rate in 1990 is associated with a 0.66 lower predicted income percentile for children born to parents at the 25th percentile for income over twenty years later. These findings are consistent with pretrial detention having an adverse impact on young children who may be the dependents of individuals affected by the pretrial system, although we caution that we are unable to control for many potential differences between low- and high-detention areas.

Our complementary pieces of evidence suggest that reducing the scope of the pretrial system, such as through the elimination of money bail, is likely to generate significant economic returns for both directly affected individuals and the communities they live in. Related reforms such as cite and release policies in lieu of arrests are also promising ways to limit the number of individuals at risk of pretrial detention. In contrast, reforms that limit the ability of employers to ask about criminal records, such as ban-the-box policies, may come too late to mitigate the economic harms associated with detention given the likely cumulative and scarring effects of pretrial detention. These later-stage reforms may also yield unintentional consequences for minority individuals (Doleac and Hansen 2020; Agan and Starr 2018).

The remainder of the paper proceeds as follows. In section I, we provide a brief background of the pretrial system, who it affects, and why it may have an impact on both individual and aggregate economic outcomes. Section II provides some descriptive statistics on pretrial detention and economic outcomes. Section III describes our evidence on the effects of pretrial detention on detained individuals at both the individual and aggregate levels. Section IV concludes and discusses areas of future work.

Figure 1. Criminal Justice Process from Arrest to Sentencing



Source: Authors.

I. A Brief Overview of the Pretrial System

The pretrial system serves as one of the first points of entry into the US criminal justice system. Figure 1 presents a simplified diagram of the various key interactions that can occur within the criminal justice system. In most jurisdictions, individuals appear for their first court appearance approximately 24–48 hours after arrest and booking by law enforcement. Figure 1 highlights that individuals can be detained in the pretrial system without eventually being either convicted of a crime or incarcerated post-conviction. In Philadelphia and Miami, for example, over 40 percent of detained individuals are not convicted of a crime and nearly 70 percent of detained individuals serve no additional incarceration (Dobbie, Goldin, and Yang 2018).

I.A. The Purpose of the Pretrial System

The US pretrial system is meant to allow all but the most dangerous criminal suspects to be released from custody while they await trial. Under the Eighth Amendment of the US Constitution, "excessive bail shall not be required, nor excessive fines imposed, nor cruel and unusual punishments inflicted." Applied to the pretrial system, the importance of release is grounded in the presumption of innocence, an axiomatic and elementary right designed to protect defendants before any finding of guilt. The pretrial system reflects the notion that pretrial detention should be used only in limited circumstances and is not deemed appropriate simply because a defendant may be guilty of the alleged crime.

The main objective of the pretrial system is to guarantee appearance at court. The federal system, along with at least forty states, also considers public and community safety explicitly as part of the release or detention

decision (Baughman and McIntyre 2012). Today, these competing objectives are embodied in the standards of the American Bar Association, which state that the judicial decision of whether to release or detain a defendant requires judges to "strike an appropriate balance" between the competing societal interests of individual liberty, court appearance, and public safety (American Bar Association 2007, 29–30).

In most jurisdictions, bail judges are granted substantial discretion when making decisions about pretrial release. These bail judges generally may consider factors such as the nature of the alleged offense, the weight of the evidence against the defendant, any record of prior flight or bail violations, and the financial ability of the defendant to pay bail. Today, with the rise in pretrial risk assessment instruments like the Arnold Ventures Public Safety Assessment (PSA), judges may also rely on a pretrial recommendation based on a risk score.

Bail judges have several options in setting conditions for pretrial release after assessing an individual's flight or public safety risk. For defendants who pose a minimal risk of flight or danger, the judge may simply release the defendant—known as release on recognizance (ROR) or personal recognizance—in which the defendant promises to return for all court proceedings and abide by all other release conditions. Defendants may also be released subject to some form of nonmonetary conditions, sometimes known as conditional release, when a judge determines that certain conditions are necessary to prevent flight or harm to the public. These conditions can include regular reporting to a pretrial services officer, drug treatment or testing, no-victim-contact orders, and even more intensive measures such as electronic monitoring or home confinement.

A judge may also impose monetary bail (cash bail or bond), which generally requires defendants to post either the full bail amount or some fraction of the bail amount to secure release. Since the twentieth century, the primary means of ensuring pretrial compliance in the United States has been the use of monetary bail as defendants are able to largely recoup their payments if they comply with all release conditions. In many jurisdictions, those who do not have the required deposit in cash can borrow this amount from commercial bail bondsmen or sureties. These agents will often accept cars, houses, jewelry, or other forms of collateral and generally charge a nonrefundable fee of 10 to 20 percent of the bail amount for their services. Another common type of monetary bail is an unsecured bond, which requires the defendant to promise to pay a certain amount of money if they do not return to court but does not require an upfront payment to secure release. If the defendant fails to appear or commits a new

crime (broadly known as pretrial misconduct), either the defendant or the bail bondsman is theoretically liable for the full value of the bail amount and forfeits any amount already paid. The amount of monetary bail may be unilaterally determined by the judge or prespecified in a bail schedule, which determines bail amounts for each type of offense or grade of offense. For example, a bail schedule might specify that a level 1 felony is associated with a $50,000 bail amount. Bail schedules are regularly used in some parts of the country, such as California and Texas, although a judge often has the discretion to reduce (or increase) the amount.

Finally, for the most serious crimes, the bail judge may require that the defendant be detained pending trial by denying bail altogether. In many jurisdictions, denial of bail is often mandatory in first- or second-degree murder cases. However, it can also be imposed for other crimes, such as domestic violence, when the bail judge finds that no set of conditions for release will guarantee appearance or protect the community from the threat of harm posed by the defendant.

In many parts of the country, determinations of bail and conditions of release are decided in short hearings that last anywhere from ten seconds to a few minutes. Defendants are often videoconferenced in from the jail, and a judge reviews the case and criminal history of the defendant, sometimes asking the defendant a few questions, before imposing conditions for pretrial release. If present, a prosecutor or defense attorney may also present their recommendations for bail.

I.B. Who Is Affected by Pretrial Detention

Decisions made at the pretrial stage affect more than 10 million individuals each year who are arrested for an offense in the United States. In some parts of the country, defendants detained before trial represent over 75 percent of all jail inmates.[2] A large contributor to the high rate of pretrial detention in the United States is the increasing use of monetary (or cash) bail, and the corresponding decreasing use of ROR over the past several decades. For example, between 1990 and 2009, in seventy-five of the most populous US counties, the share of defendants assigned monetary bail exceeded 40 percent in 2009, an 11 percentage point increase from 1990 (Reaves 2013). Over this same time period and sample, the fraction of defendants released on their own recognizance decreased by about

2. Pretrial detainees are housed in local jails. Other jail inmates include individuals serving relatively short post-conviction sentences. Prison inmates are individuals serving longer post-conviction sentences.

13 percentage points, with only 14 percent of defendants being released on their own recognizance in 2009 (Reaves 2013).

Today, the widespread use of monetary bail in many jurisdictions has resulted in high pretrial detention rates. In 2009, among felony defendants assigned monetary bail in the seventy-five largest counties, 46.9 percent were detained for the entirety of their case. Detention rates are high even when defendants are assigned relatively small monetary bail amounts. In New York City, for example, an estimated 46 percent of all misdemeanor defendants and 30 percent of all felony defendants were detained before trial in 2013 because they were unable or unwilling to post bail set at $500 or less (New York City Criminal Justice Agency 2014). Most available evidence suggests that defendants often have low earnings and rates of employment, suggesting that detention for even relatively small amounts may be due to inability to pay bail, either directly or through a bail bondsman. For example, among individuals detained in Philadelphia (for 2007–2014) and Miami (for 2006–2014), only 32 percent were employed in the year prior to arrest, only 77.2 percent had any income, and the average annual income was $4,524 (Dobbie, Goldin, and Yang 2018).

The high rate of pretrial detention and its disparate prevalence across demographic groups has contributed to concerns regarding the effectiveness of the current bail system. Critics of the current system argue that pretrial detention generates substantial costs to detainees that far outweigh the benefits to society (Pinto 2015). Critics argue, for example, that pretrial detention increases the risk of wrongful conviction by pressuring defendants to accept plea bargains to get out of jail. Pretrial detention and excessive bail conditions may also generate collateral consequences outside of the criminal justice system by disrupting defendants' lives, putting jobs, housing, and child custody at risk. These critics also argue that many jurisdictions set bail without adequate consideration for the defendant's ability to pay. As a result, they argue that pretrial detention is determined by a defendant's wealth, not their risk to the community, which reduces the current system's effectiveness and simultaneously exacerbates socioeconomic disparities. These concerns led the Department of Justice to conclude that the pretrial systems in many jurisdictions "are not only unconstitutional, but . . . also constitute bad public policy" (US Department of Justice 2016, 13).

A second concern of the current system deals with the presence of large and persistent disparities in the treatment of seemingly similar defendants, in particular by race and ethnicity. There are significant disparities in bail conditions and pretrial detention among defendants who are similar across

legally relevant dimensions in most large US counties, contributing to the overrepresentation of certain demographic groups in the pretrial system. Controlling for observable and legally relevant charge and defendant characteristics, nationally representative data on felony defendants in state courts show that, on average, Black and Hispanic defendants are substantially more likely to be detained before trial compared to observably similar white defendants (Demuth 2003; McIntyre and Baughman 2013). Disparate rates of pretrial detention are likely due to the fact that Black and Hispanic defendants are generally more likely to be assigned monetary bail and higher monetary bail amounts, compared to observably similar white defendants (Demuth 2003; Demuth and Steffensmeier 2004; Schlesinger 2005; Arnold, Dobbie, and Yang 2018). There is also substantial heterogeneity in the size of the racial gap in detention rates even after accounting for the relevant case and defendant characteristics. Harris County in Texas, for example, is 34 percent more likely to detain Black defendants compared to white defendants with the same observable characteristics, while Baltimore County in Maryland is 1 percent less likely to detain Black defendants compared to white defendants (Dobbie and Yang 2019). In recent quasi-experimental work that exploits the release tendencies of quasi-randomly assigned bail judges, researchers have found that these racial disparities can be attributed to substantial statistical discrimination and forms of racial bias (Arnold, Dobbie, and Yang 2018; Arnold, Dobbie, and Hull 2020).

Based in part on the above concerns, the pretrial systems in many jurisdictions are in flux as there is significant public support for reforming the pretrial system in the United States. In recent years, cities across the country have implemented widely supported changes to their pretrial systems. Sometimes these changes arise due to lawsuits challenging the constitutionality of money bail and bail schedules. For example, in April 2017, a federal judge in Houston issued a preliminary injunction on the current misdemeanor bail system in Harris County, Texas. Similar lawsuits are under way in many other large cities across the country. Several jurisdictions have voluntarily explored alternatives to pretrial detention, such as electronic or in-person monitoring for low-risk defendants. New York, for example, has earmarked substantial funds to supervise low-risk defendants instead of requiring them to post bail or face pretrial detention. Risk assessment instruments, such as the Arnold Ventures PSA, have been adopted by more than thirty-nine jurisdictions across the country, based on the promise of being able to more accurately predict offender risk of danger or flight. Other jurisdictions, such as New Mexico and New Jersey, have enacted large-scale reforms to their systems, effectively eliminating cash bail. In

addition, a wave of community-based efforts to change the current pretrial system has also swept the country to counteract the effects of cash bail, with charitable bail organizations like the Bail Project posting bail on behalf of eligible individuals.

I.C. Why the Pretrial System May Affect Economic Outcomes

There are two main channels through which the pretrial system can affect economic outcomes at the individual level. The first is through the direct effect of pretrial detention on the individuals who are actually detained. Even a short period of pretrial detention can be destabilizing for detained individuals, resulting in immediate job loss and affecting the extensive margin of employment, which can subsequently affect take-up of government benefits tied to formal sector employment (Dobbie, Goldin, and Yang 2018). Pretrial detention can also have a longer-run destabilizing effect on detained individuals even after the period of incapacitation ends through, for example, the stigma of a criminal conviction and lower future employment prospects, which can affect both labor supply and labor demand (Pager 2003; Agan and Starr 2018).

The second way that pretrial detention can have an impact on the economic outcomes of individuals is through spillovers on other individuals in the household or community at large. Anecdotal evidence suggests that the burden of additional caretaking responsibilities, as well as financial responsibilities that accompany monetary bail, often falls on the family members and friends of detained individuals. The spillover effects of pretrial detention may also extend across generations, as the majority of detained individuals are parents to children under the age of 18, some of whom may be placed into child custody as a result of pretrial detention. Estimates of just the direct effect of pretrial detention on detained defendants may therefore understate the effect of pretrial detention on individuals more generally.

The negative effects of the pretrial system on individual-level outcomes (both direct and spillover) can also translate into worse aggregate macroeconomic outcomes, such as employment rates. Pretrial detention removes detained individuals from the labor market during the period of detention through a short-run incapacitation effect. In a standard neoclassical model, this short-run reduction in labor supply will lower aggregate employment unless labor demand is perfectly inelastic. In practice, however, our measure of employment—the employment-to-population ratio—may be unaffected by this short-run effect given that detained individuals are removed from both the numerator (employed individuals) and denominator (noninstitutionalized civilian population). The evidence from section III.A

also suggests that the negative effects of pretrial detention on individual economic outcomes are not driven by these short-run incapacitation effects, so we do not primarily focus on this channel in our work.

Pretrial detention can also have an adverse impact on aggregate macroeconomic outcomes through the sustained and cumulative scarring effect of pretrial detention on individuals, affecting decisions to invest in human capital which in turn affect job productivity. Smith (2021) documents stories of individuals who made decisions that could have an impact on human capital accumulation following short stints in pretrial detention, such as dropping classes that were needed to receive certification for certain jobs or taking on odd jobs in between school and formal employment to pay off court-related debt in the form of fines and fees. Smith also documents stories of frequent assault and trauma during the period of detention, resulting in difficulties assimilating back to school and work after release. There is also substantial evidence from both qualitative and quantitative studies that pretrial detention causally increases future criminal legal involvement (Smith 2021; Heaton, Mayson, and Stevenson 2017; Dobbie, Goldin, and Yang 2018), with Smith (2021) and Smith and Broege (2020) arguing that pretrial detention puts people on a fundamentally different criminal justice trajectory, further leading to skill depreciation, reduced human capital investment, and reduced incentives to search for work. These sustained and cumulative scarring effects can thus reduce the labor supply of detained individuals and other individuals in the household or community at large, generating reductions in aggregate employment based on the share of affected individuals relative to the overall population. The scarring effects that lead to skill depreciation or reduced investment can also lower the labor demand for these individuals, again generating reductions in aggregate employment, particularly if there are frictions like search costs.

The goal of this paper is to produce micro-level estimates of the direct effect of pretrial detention on individuals who are actually detained, as well as macro-level aggregate estimates that combine the direct and spillover effects of pretrial detention through the channels discussed in this section.

II. National Trends in Pretrial Detention and Economic Outcomes in the United States

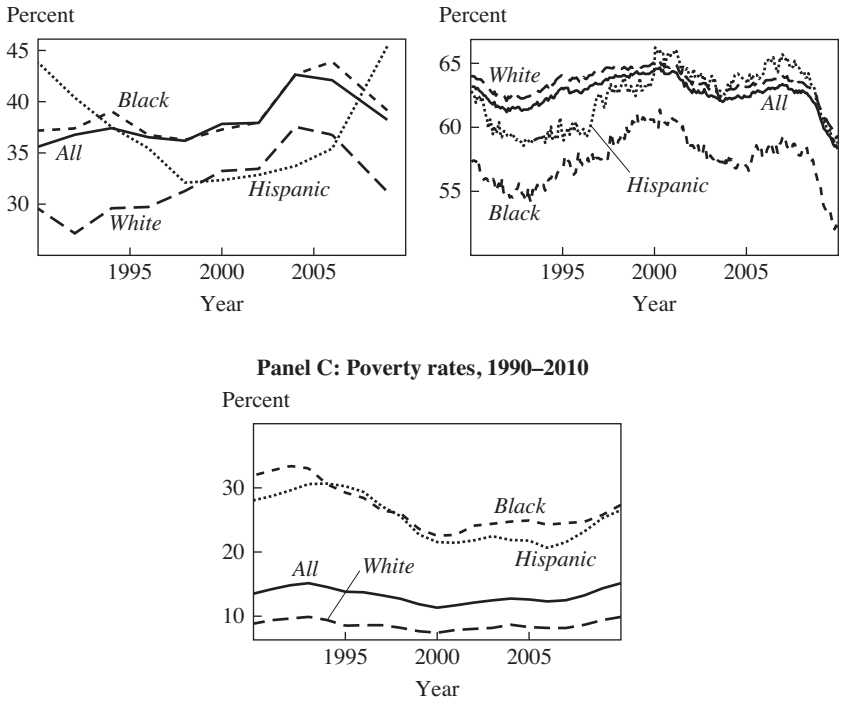
We begin by documenting the characteristics and trends of felony defendants entering the US pretrial system from 1990 to 2009 in the seventy-five largest US counties. Our data come from the Bureau of Justice Statistics' State Court Processing Statistics (SCPS), which are designed

as a nationally representative sample of seventy-five large urban counties and include information on over 140,000 individuals arrested for felony offenses. These seventy-five counties account for more than a third of the US population and approximately half of all reported crimes. The data track what happens to each defendant from the time of arrest to case disposition, providing detailed information on the arrest offense, defendant demographics and criminal history, and, importantly for our analysis, whether or not the individual was released or detained pretrial. The data include approximately forty of the nation's seventy-five most populous counties in even numbered years from 1990 to 2006 and 2009, as well as weights that allow researchers to calculate statistics that are representative of the full set of seventy-five populous counties.

Online appendix table A1 presents descriptive statistics on these felony defendants in the seventy-five most populous counties. Column 1 presents descriptive statistics for the full sample of defendants. The data reveal that 23 percent of felony defendants are non-Hispanic white, 8.1 percent are Hispanic white, and 46.9 percent are Black. Most felony defendants in the data are male (83.2 percent) and relatively young, with 35.8 percent age 24 and under and almost 90 percent age 44 and under. The most common lead charges are drug offenses (34.9 percent), followed by property offenses (30.8 percent), violent offenses (24.8 percent), and public order offenses (9.4 percent). There is an average of 4.6 prior arrests, 2.9 prior felony arrests, 2.6 prior convictions, and 1.2 prior felony convictions in our sample. In addition, there is an average of 0.5 past failures to appear for court appearances.

Columns 2 and 3 present corresponding descriptive statistics for felony defendants who are released and detained, respectively. In the SCPS, defendants are defined as detained if they remained in jail for the entirety of the time from arrest until case disposition. These individuals can be detained either because of denial of any bail (as can be the case for capital offenses or domestic violence) or because of inability to pay the assigned monetary (or cash) bail. Release and detention decisions are far from random. Non-Hispanic white defendants comprise a larger share of released defendants (25.2 percent) compared to their share among detained defendants (19.6 percent), while Black defendants comprise a lower share of released defendants (46.4 percent) compared to detained defendants (47.5 percent). We also see an overrepresentation of female defendants and defendants under age 24 among released defendants, as well as defendants with less substantial prior criminal histories (either in terms of arrests, convictions, or past failures to appear). For example, released defendants

Figure 2. Trends in Detention and Employment Rates



Sources: State Court Processing Statistics (SCPS); Bureau of Labor Statistics; and the Census Bureau.

Note: Panel A uses county-year level SCPS weights to report the share of arrested felony defendants detained pretrial in a representative sample of the nation's seventy-five most populous counties from 1990 to 2009. In the SCPS data, we define "white" as non-Hispanic white and "Hispanic" as Hispanic white. Race and ethnicity trends in the SCPS are presented only for defendants who are not missing race and ethnicity information. Panel B reports the employment-to-population ratio from 1990 to 2010 for the entire country, using data from the Bureau of Labor Statistics' Current Population Survey (Household Survey). The BLS data define "white" as all individuals identifying as white and "Hispanic" as all individuals identifying as Hispanic. Panel C reports the poverty rate from 1990 to 2010 for the entire country, using data from the Census Bureau's Current Population Survey (Annual Social and Economic Supplement). The Census Bureau data define "white" as non-Hispanic white and "Hispanic" as all individuals identifying as Hispanic.

have 3.8 prior arrests on average, while detained defendants have 5.9 prior arrests on average.

Panel A of figure 2 examines how detention rates have evolved from 1990 to 2009 among felony defendants in the seventy-five most populous counties, where we again define defendants as detained if they remained in jail for the entirety of the time from arrest until case disposition. The detention rate for felony defendants increased from 35.6 percent in 1990

to 42.1 percent in 2006, before falling slightly to 38.2 percent in 2009. As described in past work, the general upward trend in detention rates is due to the increasing use of monetary bail and corresponding decreasing use of ROR during the time period (Reaves 2013). The data also show that the detention rate for Black defendants was higher than the detention rate for non-Hispanic white defendants for the entire sample period. In 1990, for example, 37.2 percent of Black defendants were detained compared to 29.6 percent of non-Hispanic white defendants. In 2009, 39.2 percent of Black defendants were detained compared to 31.2 percent of non-Hispanic white defendants.

To motivate our later analysis, Panel B of figure 2 presents trends in the employment-to-population ratio for individuals by race or ethnicity during about the same time period of 1990 to 2010. These monthly data on employment rates are obtained from the Bureau of Labor Statistics. We view the employment-to-population ratio as the best comprehensive measure of labor market prospects, particularly for Black individuals given the rise in nonwork (Bayer and Charles 2018; Rodgers 2019).[3] The employment-to-population ratio increased from 62.8 percent in 1990 to 64.4 percent in 2000, before declining to 59.3 percent in 2009 during the Great Recession. Notably, there is a substantial racial gap in the employment-to-population ratio throughout the entire sample period, with a lower employment-to-population ratio among Black individuals compared to both white and Hispanic individuals. In 1990, for example, the average employment-to-population ratio was 56.7 percent for Black individuals and 63.7 percent for white individuals. In 2009, the average employment-to-population ratio was 53.2 percent for Black individuals compared to 60.2 percent for white individuals. These racial gaps are also present among other metrics of employment, such as the labor force participation rate and unemployment rate (Neal and Rick 2014; Bayer and Charles 2018).

Panel C of figure 2 similarly presents trends in poverty rates for individuals by race or ethnicity during the same time period. These data are obtained from the Census Bureau's Current Population Survey, Annual Social and Economic Supplement. Panel C shows that aggregate poverty

3. The employment-to-population ratio is calculated as the number of employed people divided by the civilian noninstitutionalized population. The civilian noninstitutionalized population does not include people confined to prisons or jails, and as such does not include those who are in pretrial detention in jails at the time of the survey. However, given the relatively short-term nature of pretrial detention for most defendants, we would expect many individuals who have experienced pretrial detention to be counted in the numerator and denominator of the employment-to-population ratio.

rates remained relatively consistent over the two decades, decreasing slightly from the mid-1990s to 2000 and increasing slightly thereafter. However, this aggregate trend masks substantial variation by race or ethnicity. Poverty rates for Black and Hispanic individuals declined sharply from 1990 to 2000. For example, the average poverty rate for Black individuals was 31.9 percent in 1990 and 25.8 percent in 2009. After 2000, the poverty rates for these two groups increased slightly, matching the aggregate trend. There is a clear racial gap throughout the entire period, with white poverty rates at least 12.5 percent lower than Black and Hispanic poverty rates at every point in time.

Taken together, figure 2 reveals substantial racial gaps in pretrial detention, employment-to-population ratios, and poverty rates. We also see similar broad trends in key time periods such as 2000 to 2006 (before the onset of the Great Recession), where pretrial detention rates rose sharply while employment-to-population ratios declined and poverty rates increased. In addition, Black-white gaps in pretrial detention, the employment-to-population ratio, and poverty rates all widened between 2000 and 2006. The common trends in pretrial detention and economic well-being, both overall and by race or ethnicity, raise the question of whether there is a causal relationship between pretrial detention and economic outcomes. We now turn to this question using a variety of complementary data sources and approaches.

III. The Economic Consequences of Pretrial Detention

In this section, we describe our results showing that there are real and substantial economic costs to pretrial detention, both at the individual and aggregate level. We begin by describing quasi-experimental estimates from recent work that measure the direct effects of pretrial detention on detained individuals' outcomes. We then examine the relationship between pretrial detention and aggregate macroeconomic measures of economic well-being, which also include potential spillover effects on other individuals in the household or community at large. Finally, we more tentatively explore the relationship between pretrial detention and intergenerational mobility among children.

III.A. Pretrial Detention and Individual-Level Economic Outcomes

Causal evidence on pretrial detention and the individual-level labor market comes primarily from Dobbie, Goldin, and Yang (2018), who leverage large-scale administrative data on criminal defendants to estimate

the impact of pretrial detention on a range of individual-level outcomes that capture many important costs of pretrial detention. The authors exploit plausibly exogenous variation in pretrial decisions from the quasi-random assignment of cases to bail judges who vary in their detention or release propensities. The so-called judge-IV empirical design utilized in this paper recovers the direct effect of pretrial detention for individuals at the margin of detention, meaning cases in which bail judges disagree on the appropriate bail conditions.[4]

To measure economic outcomes, Dobbie, Goldin, and Yang (2018) link data on criminal defendants to administrative tax records from the Internal Revenue Service (IRS) to examine the effects of pretrial detention on post-trial economic outcomes such as formal sector employment, individual and household income, and the take-up of government benefits. The IRS data include every individual who has ever acquired a social security number, including those who are institutionalized. Information on formal sector earnings and employment comes either from annual W-2s issued by employers or from tax returns filed by individual taxpayers. Individuals with no W-2s or self-reported income in any particular year are assumed to have had no earnings in that year. The IRS data also include information on UI from information returns filed with the IRS by state UI agencies and information on the EITC claimed by taxpayers on their return. For additional details on the IRS data and how the authors measure formal sector employment, individual earnings, and total household earnings, see Dobbie, Goldin, and Yang (2018). The authors were able to link over 300,000 criminal defendants (both misdemeanor and felony) arrested in Miami and Philadelphia from 2007 to 2014 to administrative IRS data.

The authors find that baseline earnings and formal labor market attachment are very low among arrested individuals in Miami and Philadelphia. Among defendants who are detained for at least three days pretrial, only 32 percent are employed in the year prior to arrest, 77.2 percent have any income, and the average annual income is $4,524. Among defendants released within three days, 42.3 percent are employed in the year prior to

4. The judge-IV design requires an assumption of first-stage monotonicity (Imbens and Angrist 1994; Heckman and Vytlacil 2005), which imposes a strong restriction on how judges choose which defendants to release before trial. This first-stage monotonicity assumption has received recent scrutiny both in general (Mogstad, Torgovitsky, and Walters 2019) and in the specific context of judge-IV designs (Mueller-Smith 2015; Frandsen, Lefgren, and Leslie 2019; Norris 2019; Arnold, Dobbie, and Hull 2020). Recent work argues that the monotonicity assumption is unlikely to hold exactly in judge-IV designs but that these estimates can still identify a convex combination of treatment effects under a weaker assumption of average monotonicity (Frandsen, Lefgren, and Leslie 2019).

Table 1. Pretrial Detention and Individual Outcomes

	Detained mean *(1)*	*2SLS estimates* *(2)*	*NPV estimates* *(3)*
Panel A: Binary outcomes			
Any formal sector earnings	0.378	−0.094	—
	(0.485)	(0.057)	
Any unemployment insurance	0.064	−0.013	—
	(0.246)	(0.033)	
Any earned income tax credit	0.233	−0.105	—
	(0.423)	(0.049)	
Panel B: Outcomes in dollars			
Formal sector earnings	5,887	−948	−18,961
	(15,897)	(1,128)	
Unemployment insurance	245	−293	−5,860
	(1,335)	(193)	
Earned income tax credit	357	−209	−4,180
	(998)	(127)	
Observations	144,290	334,943	—

Source: Dobbie, Goldin, and Yang (2018).

Note: Column 1 reports the mean outcome for detained defendants in the sample. Column 2 reports two-stage least squares (2SLS) estimates that instrument for pretrial detention using a judge leniency measure that is estimated using data from other cases assigned to a bail judge in the same year, controlling for court-by-time fixed effects and baseline controls. Column 3 reports the net present value of each change at a 3 percent discount rate, under the assumption that the percentage point gain for each outcome remains constant over the working life cycle. The standard deviations of each outcome for detained defendants are reported in parentheses in column 1 and robust standard errors two-way clustered at the individual and judge level are reported in parentheses in column 2. All outcomes are measured three to four years after an individual's arrest using administrative tax data.

arrest, 81.4 percent have any income, and the average annual income is $7,223. These descriptive statistics indicate that arrested individuals are likely to be different from the general working-age population.[5]

Moving to two-stage least squares estimates, Dobbie, Goldin, and Yang (2018) find that pretrial detention causally decreases both attachment to the formal labor market and the receipt of employment- and tax-related government benefits in their sample of Miami and Philadelphia cases. Table 1 summarizes the main results based on these linked data from Dobbie, Goldin, and Yang (2018). As reported in panel A of table 1, pretrial detention decreases the probability of employment in the formal labor market three to four years after the bail hearing by 9.4 percentage points in this sample, a 24.9 percent decrease from the mean. The authors find evidence that pretrial detention primarily affects earnings at the extreme low end of the income distribution, with little discernible effects at other points

5. Grogger (1995) finds similar patterns of substantial joblessness in a sample of adult individuals in California arrested between 1973 and early 1987 matched to UI records.

of the distribution. In terms of UI receipt and EITC receipt—measures of formal sector engagement that are particularly welfare-relevant in our low-income population—pretrial detention also decreases the probability that the defendant takes up any UI benefits within three to four years after case disposition by 1.3 percentage points, a 20.3 percent decrease, and decreases the take-up of EITC benefits by 10.5 percentage points over the same time period, a 45.1 percent decrease.

Panel B of table 1 presents results on outcomes in dollars. Pretrial detention reduces formal sector earnings by $948 per year over the same time period, a 16.1 percent decrease from the mean. In terms of UI receipt and EITC receipt, the authors find that three to four years after arrest, pretrial detention decreases UI benefits by $293 and EITC benefits by $209 per year, 119.6 and 58.5 percent decreases from the mean, respectively. All of the estimated effects are larger among individuals with no prior offenses in the past year. These results are consistent with the stigma of a criminal conviction lowering defendants' prospects in the formal labor market (Pager 2003; Agan and Starr 2018) and reduced labor supply and human capital accumulation (Smith 2021; Smith and Broege 2020), which in turn limits defendants' eligibility for employment-related benefits like UI and EITC.

The findings from the recent quasi-experimental empirical literature thus suggest that pretrial detention imposes substantial economic costs to individual defendants at the margin, reducing formal sector employment at the extensive margin. The findings from Dobbie, Goldin, and Yang (2018) also suggest that the long-run effects on employment are unlikely to be primarily driven by a short-run incapacitation effect (e.g., losing one's job while being detained pretrial), given that employment is lower long after the period of pretrial detention. The longer-run effect of pretrial detention on formal sector employment is likely driven by lower labor supply by detained individuals or lower labor market demand for individuals with criminal records, as discussed above.

Applying the two-stage least squares estimates from Dobbie, Goldin, and Yang (2018), we can use back-of-the-envelope calculations to explore the amount that each marginally detained defendant loses in income over the course of the working life cycle (see column 3 of table 1). Recall that the marginal detained defendant earns roughly $948 less per year and has $293 less in UI income and $209 less in EITC income, for a total average annual income loss of $1,450, 10.1 percent of mean earnings in the sample. Following Chetty, Friedman, and Rockoff (2014), we assume that the percentage loss in earnings remains constant over the working life cycle and discount annual earnings at a 3 percent discount rate back to age 34,

the mean age in the sample. Under these assumptions, the marginal detained defendant loses $29,000 in income over a lifetime relative to the marginal released defendant, with almost $19,000 of the lost income due to reduced formal sector earnings—confirming the substantial economic costs to detained defendants at the margin suggested in the recent literature.

POLICY IMPLICATIONS We can use the quasi-experimental estimates from Dobbie, Goldin, and Yang (2018) to evaluate two types of policy reforms. The first policy we consider is shifting the rates in high-detention counties such as Harris County, Texas, with a detention rate of 65 percent, to match those in low-detention counties such as Broward County, Florida, with a detention rate of 22.2 percent. The second policy we consider is the elimination of cash bail, which the available evidence suggests would lower the pretrial detention rate to anywhere from 3 to 10 percent.[6] We note that these simulations can only yield approximate estimates given the imprecision of some quasi-experimental estimates from Dobbie, Goldin, and Yang (2018).

For our first policy simulation, we estimate the number of affected individuals using data from the FBI's Uniform Crime Reporting Program, which shows that there were 346,140 offenses recorded in cities in Harris County in 2009.[7] Reducing Harris County's detention rate from 65 percent to 22.2 percent to match that of Broward County thus implies that up to 148,147 people would not have been detained under our first policy counterfactual. Since the quasi-experimental estimates from Dobbie, Goldin, and Yang (2018) suggest that each newly released individual gains $29,000 in income over their lifetime, this means that reducing Harris County's detention rate from 65 percent to 22.2 percent would increase aggregate income in the county by approximately $4.30 billion over the lifetime of those released defendants.

For our second policy simulation, we estimate the number of affected individuals using the 10.08 million arrests for all offenses in the United States. The current pretrial detention rate for felony defendants is 37.71 percent, according to the 2009 SCPS. If we assume that this was the rate of pretrial detention for all arrested individuals, then this implies that if

6. For example, the outright detention for felony offenses in the 2009 SCPS is 3.9 percent of individuals, while pretrial detention rates were as low as 6 percent following the eradication of money bail in New Jersey. We also observe detention rates as low as 10 percent in parts of country that have traditionally not relied on money bail such as the District of Columbia.

7. The 2009 data from the FBI's Uniform Crime Reporting Program were released as part of the report *Crime in the United States, 2009*. These data were accessed on March 9, 2021; Crime in the United States, https://www2.fbi.gov/ucr/cius2009/documents/index.html.

money bail was eliminated and the pretrial detention rate fell to 10 percent, 2.79 million people would no longer be detained.[8] Again applying the quasi-experimental estimates from Dobbie, Goldin, and Yang (2018), this means that eliminating money bail in the United States could increase aggregate income by up to $80.91 billion per year, assuming the same number of people are no longer detained each year. We note that this is an upper bound on the change in aggregate income as equilibrium effects would tend to dampen the income effects of an increase in labor supply associated with elimination of money bail.

One important caveat is that the quasi-experimental estimates from Dobbie, Goldin, and Yang (2018) do not capture potential spillover effects of pretrial detention on other individuals. These quasi-experimental estimates are also based on the approximately 13 percent of defendants at the margin of release, not the average defendant who might be released. The estimates based on these marginal defendants may not be applicable to the average defendant who would be affected by large-scale policy reforms. We thus tentatively explore whether pretrial detention has an impact on aggregate economic measures using policy-relevant variation in detention rates and an empirical design that also captures potential spillover effects.

III.B. County-Level Detention Rates and Economic Outcomes

We explore the aggregate effects of pretrial detention inclusive of spillover effects by comparing changes in county-level poverty and employment-to-population rates to changes in county-level pretrial detention rates. We take this approach because there is substantial heterogeneity across counties in their change in pretrial detention rates from 2000 to 2009, the latest year of available SCPS data. For example, between 2000 and 2009, counties like Wayne, Michigan, and Franklin, Ohio, experienced over 15 percentage point increases in detention rates compared to counties like Broward, Florida, and Miami-Dade, Florida, which experienced over 10 percentage point decreases in detention rates.

Exploiting this time series variation, we present scatterplots and regression estimates of the following county-level specification:

$$\Delta Y_{c,2000-2010} = \alpha + \beta_1 \Delta Detention_{c,2000-2009} + \beta_2 X_{c,2000} + \varepsilon_c$$

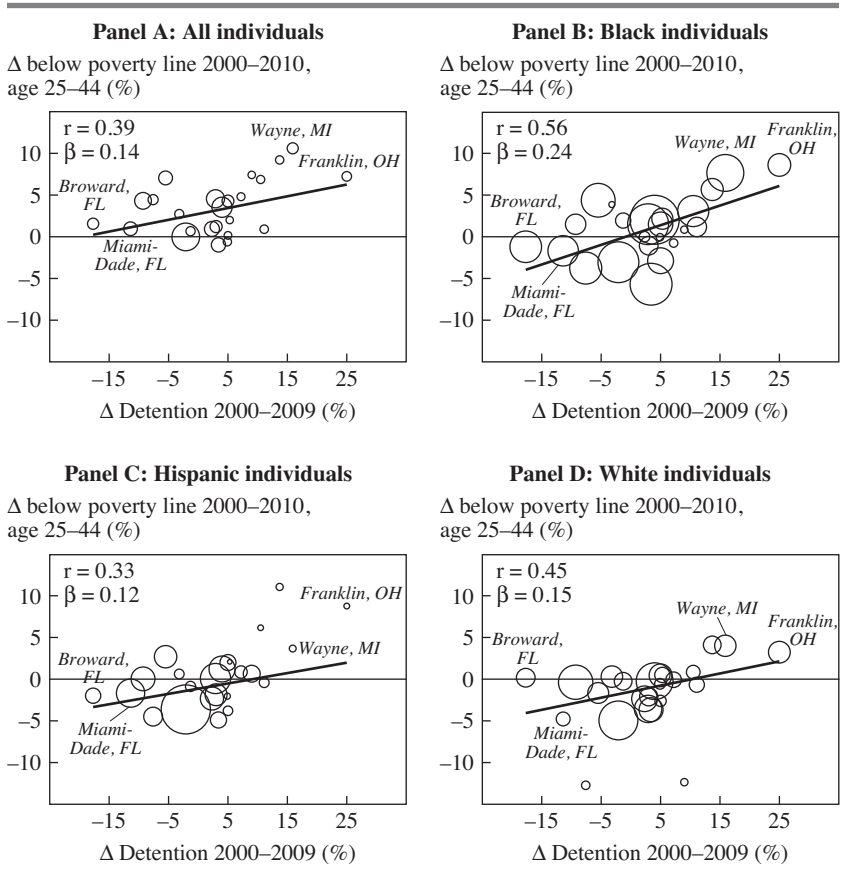
8. We note that the number of people no longer detained would likely be lower than 2.79 million given that the total number of arrests per year includes arrests for repeat offenders. This number is also a rough approximation given that we do not know the pretrial detention rate for individuals charged with misdemeanor offenses, which could be different from the rate reported in the SCPS data.

where $\Delta Y_{c,2000-2010}$ represents the 2000 to 2010 change in economic outcomes in county c (in percentage points), $\Delta Detention_{c,2000-2009}$ represents the 2000 to 2009 change in detention rates in county c (in percentage points), and $X_{c,2000}$ represents a vector of baseline county-level covariates. All results are weighted by the county-level race- and age-specific population.

Before proceeding to our estimates, we note that our county-level estimates should be interpreted with an abundance of caution. For various reasons, these county-level specifications should not be interpreted as precise or causal estimates of the relationship between changes in pretrial detention rates and changes in economic outcomes. For one, we are only able to explore these relationships among twenty-four counties in the SCPS (the best available nationally representative data) for whom we can observe detention rates in both 2000 and 2009. Thus, estimates are very imprecise. Second, we utilize county-level changes in detention rates that are likely endogenous. As such, there may be other county-level changes correlated with changes in pretrial detention rates, such that one should be cautious in interpreting β_1 as a causal effect. Nevertheless, even in the absence of convincing quasi-experimental variation in aggregate pretrial detention rates, we feel that these estimates can provide a useful first step to policymakers and researchers in shedding light on whether individual-level effects in the quasi-experimental literature may translate to aggregate economic outcomes and on the potential scope of spillover effects.

CHANGES IN DETENTION RATES AND POVERTY RATES We begin by examining the relationship between changes in detention rates and changes in poverty rates. In figure 3, we present correlations between the 2000 to 2009 change in pretrial detention rates ($\Delta Detention$) and the 2000 to 2010 change in county-level poverty rates ($\Delta Poverty$), defined as the share of the population below the poverty line. County-level poverty rates are measured using the 2000 Decennial Census and the 2006–2010 American Community Survey five-year estimates. These results show whether counties that experienced greater increases in their pretrial detention rates exhibited different changes in their poverty rates. In all results below, we present correlations and regression estimates of $\Delta Poverty$ and $\Delta Detention$ for prime working-age individuals between the ages of 25 to 44, Black individuals age 25 to 44, Hispanic individuals age 25 to 44, and non-Hispanic white individuals age 25 to 44. We focus on these demographic groups given that the pretrial system is overrepresented by Black defendants and defendants under the age of 45. The size of the circles represents the age- and race-specific county population in 2000. We explore these relationships among twenty-four counties in the SCPS for which we can observe detention rates in

Figure 3. Changes in County Detention and Poverty Rates



Sources: State Court Processing Statistics (SCPS); 2000 Decennial Census; and the American Community Survey.

Note: The change in county detention rates from 2000 to 2009 is measured using the change in the share of arrested felony defendants detained pretrial in the SCPS in the twenty-four counties with data in both 2000 and 2009. The change in county poverty rates from 2000 to 2010 is measured using the change in the poverty rate between the 2000 Decennial Census and the American Community Survey five-year estimates 2006–2010 for individuals age 25–44 in the same twenty-four counties. Correlations and best-fit regression lines are weighted using the applicable race- and age-specific county population total in 2000 as reported in the 2000 Decennial Census. The size of the circles represents the age- and race-specific county population in 2000. See table 2 for additional regression estimates and standard errors.

both 2000 and 2009. Panels A and B of online appendix table A2 present summary statistics for these changes, both unweighted and weighted by the race- and age-specific county population.

Figure 3 reveals that among individuals age 25 to 44, the correlation between Δ*Poverty* and Δ*Detention* is 0.39, suggesting that counties with larger increases in pretrial detention rates also experienced larger increases in poverty rates. For example, two counties with the largest increases in pretrial detention rates over the 2000 to 2009 time period—Wayne, Michigan, and Franklin, Ohio—experienced a growth in the poverty rate from 2000 to 2010 for individuals age 25 to 44 of 10.61 percentage points and 7.24 percentage points, respectively. In contrast, counties with large decreases in pretrial detention rates over the time period, such as Broward, Florida, and Miami-Dade, Florida, experienced a growth in the poverty rate from 2000 to 2010 for individuals age 25 to 44 of only 1.56 percentage points and 0.96 percentage points, respectively.

Consistent with these correlation estimates, a linear regression of Δ*Poverty* and Δ*Detention* with no baseline controls yields a regression coefficient of 0.14, implying that a 10 percentage point increase in the detention rate between 2000 and 2009 is associated with a 1.41 percentage point increase in the poverty rate of prime working-age individuals between 2000 and 2010. This regression coefficient similarly implies that a one standard deviation increase in the change in the detention rate between 2000 and 2009 (9.36 percentage points) is associated with a 1.32 percentage point increase in the change in the poverty rate over a similar time period. Among Black individuals age 25 to 44, the correlation coefficient between Δ*Poverty* and Δ*Detention* is 0.56 with a linear regression coefficient of 0.24, implying that a one standard deviation increase in the change in the detention rate between 2000 and 2009 is associated with a 2.21 percentage point increase in the change in the poverty rate of Black individuals age 25 to 44 between 2000 and 2010. Among Hispanic individuals age 25 to 44, the correlation coefficient is 0.33 with a linear regression coefficient of 0.13, implying that a one standard deviation increase in the change in the detention rate between 2000 and 2009 is associated with a 1.17 percentage point increase in the change in the poverty rate of Hispanic individuals age 25 to 44 over the time period. Finally, for non-Hispanic white individuals, the correlation coefficient between Δ*Poverty* and Δ*Detention* is 0.45 with a linear regression coefficient of 0.15, implying that a one standard deviation increase in the change in the detention rate between 2000 and 2009 is associated with a 1.36 percentage point increase in the change in the poverty rate of non-Hispanic white individuals age 25 to 44 over the time period.

In general, these results reveal a positive relationship between detention rate changes and poverty rate changes, with the largest correlations among Black prime working-age individuals.

In panel A of table 2, we present estimates from a regression of Δ*Poverty* on Δ*Detention* with and without baseline controls in levels. County baseline controls (measured in 2000) include the mean household income, share female, share of single parents, share foreign born, the unemployment rate, EITC exposure, violent crime rates, and total crime rates. We also include the 2000 share of population with a college degree or more, the share of women in the labor force, and the log population, following Charles, Hurst, and Notowidigdo (2016). Estimates are weighted by the county-level race- and age-specific population. Even after accounting for these baseline controls, we continue to find that a one standard deviation increase in the change in the detention rate between 2000 and 2009 is associated with a 0.98 percentage point increase in the change in the poverty rate of individuals age 25 to 44, a 3.10 percentage point increase in the poverty rate of Black individuals age 25 to 44, a 1.17 percentage point increase in the poverty rate of Hispanic individuals age 25 to 44, and a 0.54 percentage point increase in the poverty rate of non-Hispanic white individuals age 25 to 44 between 2000 and 2010. These regression estimates with controls are again consistent with a positive relationship between detention rate changes and poverty rate changes, with the largest relationship for Black individuals age 25 to 44. We caution, however, that our estimates are imprecisely estimated and that the 95 percent confidence interval includes a range of estimates. For example, the lower range of a 95 percent confidence interval suggests that a one standard deviation increase in the change in the detention rate between 2000 and 2009 is associated with a 0.96 percentage point decrease in the change in the poverty rate of individuals age 25 to 44 and a 0.36 percentage point increase in the poverty rate of Black individuals age 25 to 44.

CHANGES IN DETENTION RATES AND EMPLOYMENT-TO-POPULATION RATIOS

We now turn to the employment rate, measured by the employment-to-population ratio. In figure 4, we present correlations between the 2000 to 2009 change in pretrial detention rates (Δ*Detention*) and the 2000 to 2010 change in county-level civilian employment rates (Δ*Employment*), defined as the employment-to-population ratio among the civilian population. County-level employment rates are measured using the 2000 Decennial Census and the 2006–2010 American Community Survey five-year estimates. We present correlations and regression estimates of Δ*Employment* and Δ*Detention* for prime working-age individuals between the ages of 25 to 44, and all Black individuals age 16 to 64, Hispanic individuals

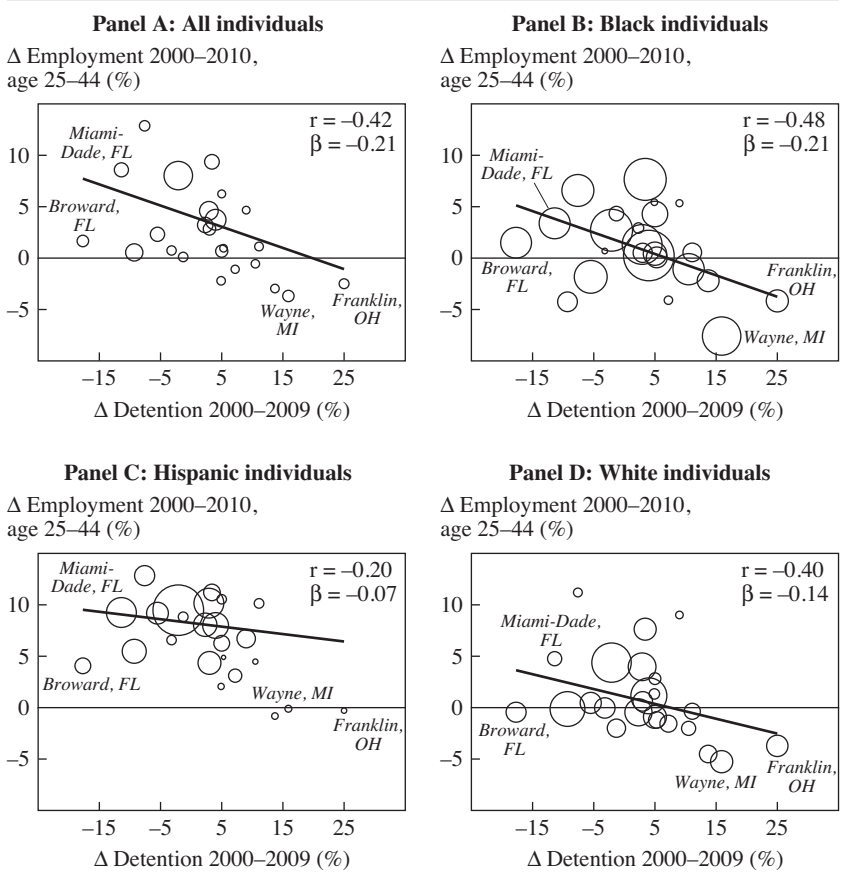
Table 2. County Detention Rates and Economic Outcomes

	All individuals		*Black individuals*		*Hispanic individuals*		*White individuals*	
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Panel A: Δ*Poverty* 2000–2009								
Δ*Detention* 2000–2009	0.141	0.105	0.236	0.331	0.125	0.125	0.146	0.058
	(0.072)	(0.106)	(0.077)	(0.149)	(0.064)	(0.074)	(0.070)	(0.078)
	24	24	24	24	24	24	24	24
Panel B: Δ*Employment* 2000–2009								
Δ*Detention* 2000–2009	–0.206	–0.115	–0.208	–0.408	–0.072	–0.056	–0.144	–0.125
	(0.109)	(0.072)	(0.103)	(0.121)	(0.088)	(0.048)	(0.075)	(0.047)
	24	24	24	24	24	24	24	24
Panel C: Mobility at the 25th percentile for parental income								
Detention rate in 1990	0.021	–0.066	–0.032	–0.055	0.006	–0.040	–0.021	–0.112
	(0.026)	(0.027)	(0.019)	(0.023)	(0.022)	(0.019)	(0.027)	(0.036)
	39	39	39	39	39	39	39	39
Panel D: Mobility at the 75th percentile for parental income								
Detention rate in 1990	–0.040	–0.070	–0.025	–0.019	–0.031	–0.045	–0.069	–0.093
	(0.017)	(0.028)	(0.023)	(0.027)	(0.024)	(0.020)	(0.021)	(0.036)
	39	39	39	39	39	39	39	39
Baseline controls	No	Yes	No	Yes	No	Yes	No	Yes

Sources: State Court Processing Statistics; 2000 Decennial Census; American Community Survey; and Opportunity Insights.

Note: Panel A reports estimates of regressions of the change in poverty rates from 2000 to 2010 (in percentage points) on the change in felony detention rates from 2000 to 2009 (in percentage points) in the twenty-four counties with detention data in both 2000 and 2009. Panel B reports estimates of regressions of the change in employment-to-population rates from 2000 to 2010 (in percentage points) on the change in felony detention rates from 2000 to 2009 (in percentage points) in the twenty-four counties with detention data in both 2000 and 2009. Panel C reports estimates of regressions of the income percentiles of children born to parents at the 25th percentile on the felony detention rate in 1990 (in percentage points) in the thirty-nine counties with detention data in 1990. Panel D reports estimates of regressions of the income percentiles of children born to parents at the 75th percentile on the felony detention rate in 1990 (in percentage points) in the thirty-nine counties with detention data in 1990. In panels A–B, the baseline controls include the mean household income, share female, share of single parents, share foreign born, the unemployment rate, EITC exposure, and violent and total crime rates. Baseline controls also include the share of population with a college degree or more, the share of women in the labor force, and log population, following Charles, Hurst, and Notowidigdo (2016). All baseline controls in panels A–B are measured in 2000. In panels C–D, the baseline controls included in the even columns are the Gini coefficient for the bottom 99 percent, high school dropout rate, the share of Black individuals, the share of single mothers, a social capital index, the violent crime rate, and the total crime rate, broadly following Chetty and others (2014). All specifications use the 2000 Census group-specific population in each county as weights, report robust standard errors in parentheses, and report the number of observations below each estimate. See figures 3–6 for additional details on the coding of each outcome measure.

Figure 4. Changes in County Detention and Employment Rates



Sources: State Court Processing Statistics (SCPS); 2000 Decennial Census; and the American Community Survey.

Note: The change in county detention rates from 2000 to 2009 is measured using the change in the share of arrested felony defendants detained pretrial in the SCPS in the twenty-four counties with data in both 2000 and 2009. The overall change in county employment-to-population rates from 2000 to 2010 is measured using the change in the employment-to-population rate between the 2000 Decennial Census and the American Community Survey five-year estimates 2006–2010 for individuals age 25–44 in the same twenty-four counties. The race-specific changes in county employment-to-population rates are measured using individuals age 16–64. Correlations and best-fit regression lines are weighted using the applicable race- and age-specific county population total in 2000 as reported in the 2000 Decennial Census. The size of the circles represents the age- and race-specific county population in 2000. See table 2 for additional regression estimates and standard errors.

age 16 to 64, and non-Hispanic white individuals age 16 to 64. Due to restrictions in the availability of public data at the county level, we are unable to further disaggregate employment rates for demographic groups by narrower age ranges (e.g., Black individuals between the ages of 25 to 44). The size of the circles represents the age- and race-specific county population in 2000. Again, we explore these relationships among twenty-four counties in the SCPS for which we can observe detention rates in both 2000 and 2009. Summary statistics on these changes are presented in panels A and C of online appendix table A2.

Figure 4 reveals that among individuals age 25 to 44, the correlation between Δ*Employment* and Δ*Detention* is −0.42. Counties with large increases in pretrial detention rates, such as Wayne, Michigan, and Franklin, Ohio, experienced employment rate decreases for individuals age 25 to 44 of 3.69 percentage points and 2.49 percentage points, respectively. In contrast, counties with large decreases in pretrial detention rates over the time period, such as Broward, Florida, and Miami-Dade, Florida, experienced employment rate increases for individuals age 25 to 44 of 1.67 percentage points and 8.59 percentage points, respectively. A linear regression of Δ*Employment* and Δ*Detention* with no baseline controls yields a regression coefficient of −0.21, implying that a 10 percentage point increase in the change in the detention rate between 2000 and 2009 is associated with a 2.06 percentage point decrease in the change in the employment rate of prime working-age individuals between 2000 and 2010. This regression coefficient similarly implies that a one standard deviation increase in the change in the detention rate between 2000 and 2009 (9.36 percentage points) is associated with a 1.93 percentage point decrease in the change in the employment rate over a similar time period.

Among Black individuals age 16 to 64, the correlation coefficient between Δ*Employment* and Δ*Detention* is −0.48 with a linear regression coefficient of −0.21. This regression coefficient implies that a one standard deviation increase in the change in the detention rate between 2000 and 2009 is associated with a 1.95 percentage point decrease in the change in the employment rate of Black individuals age 16 to 64 between 2000 and 2010. Among Hispanic individuals age 16 to 64, the correlation coefficient between Δ*Employment* and Δ*Detention* is −0.20 with a linear regression coefficient of −0.07, implying that a one standard deviation increase in the change in the detention rate between 2000 and 2009 is associated with a 0.67 percentage point decrease in the change in the employment rate of Hispanic individuals age 16 to 64. Finally, for non-Hispanic white individuals age 16 to 64, the correlation coefficient between Δ*Employment* and

Δ*Detention* is –0.40 with a linear regression coefficient of –0.14, implying that a one standard deviation increase in the change in the detention rate between 2000 and 2009 is associated with a 1.35 percentage point decrease in the change in the employment rate of non-Hispanic white individuals age 16 to 64. Thus, there is a negative association between detention rate changes and employment rate changes, particularly among Black working-age individuals.

In panel B of table 2, we present estimates from a regression of Δ*Employment* on Δ*Detention* with and without baseline controls in levels. We use the same set of baseline controls as described above in our regressions with changes in poverty rates. We weight these regressions with the relevant county-level population for each racial and age group. With baseline controls, we find that a one standard deviation increase in the change in the detention rate between 2000 and 2009 is associated with a 1.08 percentage point decrease in the change in the employment rate of individuals age 25 to 44, a 3.82 percentage point decrease in the employment rate of Black individuals age 16 to 64, a 0.52 percentage point decrease in the employment rate of Hispanic individuals age 16 to 64, and a 1.17 percentage point decrease in the employment rate of non-Hispanic white individuals age 16 to 64 between 2000 and 2010. As with our above results on poverty rates, these regression results suggest that local changes in detention rates are generally associated with changes in aggregate economic well-being as measured by employment rates, particularly for Black individuals. We again caution that our estimates are imprecisely estimated and that the 95 percent confidence interval includes a range of estimates. For example, a 95 percent confidence interval suggests that a one standard deviation increase in the change in the detention rate between 2000 and 2009 is associated with a –2.40 to 0.24 percentage point change in the employment rate of individuals age 25 to 44 and a –6.04 to –1.60 percentage point change in the employment rate of Black individuals age 16 to 64 between 2000 and 2010.

POLICY IMPLICATIONS We now return to our back-of-the-envelope calculations exploring how changes in the pretrial system could affect changes in poverty rates and employment rates. These calculations should be interpreted cautiously given the imprecise and noncausal nature of our exploratory estimates. The first policy we consider is shifting detention rates in a high Δ*Detention* county such as Franklin, Ohio, where pretrial detention rates increased 25 percentage points between 2000 and 2009, to match a low Δ*Detention* county such as Broward, Florida, where pretrial detention rates *decreased* by 17.7 percentage points between 2000 and 2009. The

second policy we consider is again the elimination of cash bail, where we assess what might have happened if all counties had reduced detention rates to only 10 percent in 2009 as compared to the actual increase in detention rates between 2000 and 2009.

Starting with the first policy simulation, if Franklin, Ohio, had reduced its Δ*Detention* to match that of Broward, Florida, over the same period, our regression estimates imply that there would have been a 4.48 (with controls) to 6.02 (without controls) percentage point decrease in Franklin's Δ*Poverty* for individuals age 25 to 44 between 2000 and 2010 on top of the actual observed change in poverty. Given that the actual poverty rate in Franklin increased by 7.24 percentage points over this time period, Franklin's counterfactual poverty rate would have increased by only 1.22 to 2.77 percentage points. The effects of this simulation are even larger among minority individuals. Among Black individuals age 25 to 44, if Franklin, Ohio, reduced its Δ*Detention* to match that of Broward, Florida, our regression estimates imply an additional 10.09 to 14.15 percentage point decrease in Franklin's Δ*Poverty* for Black individuals age 25 to 44 between 2000 and 2010 on top of the actual observed change in poverty. Since Franklin's actual poverty rate increased by 8.68 percentage points for this demographic group during this time period, the counterfactual poverty rate for Black prime working-age individuals would have decreased by 1.41 to 5.47 percentage points.

We can do the same simulation for employment rates. If Franklin, Ohio, reduced its Δ*Detention* to match that of Broward, Florida, our regression estimates imply a counterfactual employment rate increase of 2.42 to 6.31 percentage points compared to the actual decrease of 2.49 percentage points. Among working-age Black individuals, the counterfactual employment rate change for this demographic group would have been an increase of 4.72 to 13.27 percentage points as compared to the actual decrease of 4.16 percentage points.

We next consider what might happen if all counties eliminated money bail in 2009 such that detention rates were only 10 percent relative to the actual increase in detention rates between 2000 and 2009. If all counties had reduced their 2009 detention rates to 10 percent, this change (Δ*Detention*) would have represented a 31.33 percentage point decrease from the mean population weighted county detention rate in 2000 as measured in the SCPS data. In actuality, Δ*Detention* between 2000 and 2009 was 0.67 percentage points. Thus, our estimates in figure 3 imply that, on average, poverty rates for individuals age 25 to 44 in all counties would have decreased by 3.35 to 4.51 percentage points more than they actually changed if counties

eliminated money bail in 2009. Between 2000 and 2010, the weighted average poverty rate across all counties in the SCPS increased by 2.83 percentage points. Thus, if counties had eliminated money bail relative to actuality, our estimates suggest that counterfactual poverty rates over this same period would have instead decreased by 0.52 to 1.68 percentage points. For Black individuals age 25 to 44, our estimates imply that if counties had eliminated money bail relative to actuality, the counterfactual poverty rates would have shown a decrease of 6.83 to 9.88 percentage points compared to the actual increase of 0.73 percentage points.

Again, we emphasize that these simulations are based on imprecise estimates and magnitudes should not be taken literally. If we used the lower end of a 95 percent confidence interval, for example, our estimates suggest that if counties had eliminated money bail relative to actuality, counterfactual poverty rates over this same period would have instead increased by 2.84 to 6.16 percentage points for all individuals age 25 to 44 and increased by 0.09 to 1.57 percentage points for Black individuals age 25 to 44.

These changes may also yield economically large decreases in racial gaps in poverty rates given that reductions in detention have a greater differential impact on Black versus white prime working-aged individuals, although again we note that our estimates are measured with considerable noise. For example, in 2010, the population-weighted average poverty rate was 19.3 percent for Black individuals and 6.9 percent for white individuals. Therefore, the Black-white racial gap was 12.43 percentage points in 2010. If counties had eliminated money bail, our estimates predict that there would have been an additional change in the $\Delta Poverty$ of −4.66 to −5.51 percentage points for white individuals and −7.56 to −10.61 percentage points for Black individuals compared to actuality. Using the lower end of these estimates for both groups, the counterfactual 2010 white poverty rate would have been 2.2 percent and the counterfactual 2010 Black poverty rate would have been 11.7 percent. Thus, if counties eliminated money bail, the counterfactual racial gap in 2010 would have only been 9.53 percentage points, which is 23 percent smaller than the actual racial gap.

We now turn to similar policy simulations for employment rate changes. Between 2000 and 2010, the actual weighted average change in employment rates across all counties in the SCPS was a 3.94 percentage point increase for all individuals age 25 to 44 and a 1 percentage point increase for Black individuals age 16 to 64. If these counties had eliminated money bail in 2009 relative to the actual increase in detention rates over the time period, the counterfactual employment rate change across all counties

would have been a 7.63 to 10.54 percentage point increase for all individuals age 25 to 44 and a 7.67 to 14.08 percentage point increase for Black individuals age 16 to 64.

As before, we emphasize that these simulations are based on imprecise estimates and contain a wide range of possible estimates. If we used the lower end of a 95 percent confidence interval, our estimates suggest that if counties had eliminated money bail relative to actuality, counterfactual employment rates over this same period would have instead increased by 12.15 to 17.36 percentage points for all individuals age 25 to 44 and increased by 17.04 to 24.62 percentage points for Black individuals age 16 to 64. Using the upper end of a 95 percent confidence interval suggests much smaller counterfactual employment rate increases of 3.1 to 3.71 percentage points for all individuals age 25 to 44 and 4.17 to 9.42 for Black individuals age 16 to 64.

Eliminating money bail could also yield economically significant decreases in racial gaps in employment rates. If counties had reduced detention rates to 10 percent via elimination of money bail relative to reality, our estimates suggest that the counterfactual employment-to-population ratio would have been 76.87 percent for white individuals and 67.71 percent for Black individuals in 2010, implying a counterfactual racial gap in 2010 of 9.16 percentage points, 22.5 percent smaller than the actual 2010 racial gap of 11.82 percentage points. In sum, these exploratory simulations suggest that eliminating money bail may decrease poverty rates and increase employment rates, particularly among working-age Black individuals.

COMPARING DIRECT INDIVIDUAL-LEVEL AND AGGREGATE MACROECONOMIC ESTIMATES A natural question may be whether the aggregate macroeconomic effects (which include both direct and spillover effects) we find in this section can be explained by the direct individual-level estimates described above in section III.A. Here, we do a very crude comparison of these estimates for employment, noting that this exercise is highly speculative given the noncomparability of the estimates. For one, the direct individual-level estimates represent only local average treatment effects for defendants at the margin of release and may be inapplicable to inframarginal defendants affected by policy reforms. Second, the direct estimates are based on annual labor market changes, while the aggregate macroeconomic estimates capture cumulative steady-state changes over the course of a decade. Third, the aggregate macroeconomic estimates relying on county-level changes are noncausal in nature and have large standard errors. Therefore, we do not think that direct comparisons between these estimates are justified given the existing state of research. Nevertheless,

a rough calibration leads us to tentatively conclude that spillover effects are an area worthy of further exploration.

Recall that for each person no longer detained, Dobbie, Goldin, and Yang (2018) estimate that the probability of employment in the formal sector increases by 9.4 percentage points each year. Mapping this individual-level probability to an aggregate measure, such as the employment-to-population ratio, depends on what share of the relevant working-age population is at risk of pretrial detention. Based on McCauley (2017), estimates from the National Longitudinal Survey of Youth (NLSY) 1997 suggest that the cumulative arrest probability by age 28 for all respondents is 32 percent, which we use as a benchmark for the size of the at-risk population. If we assumed a 100 percentage point reduction in the detention rate for this at-risk population (effectively going from universal to no detention), the estimates from Dobbie, Goldin, and Yang (2018) imply that the aggregate employment rate would correspondingly rise by up to 3.01 percentage points (9.4 percentage points times 0.32) through direct effects. Turning to aggregate estimates, our county-level estimates of the association between changes in detention rates and changes in employment rates with baseline controls imply that a 100 percentage point reduction in the detention rate would increase the prime working-age employment rate by 11.5 percentage points (see panel B of table 2). However, given the large standard errors, within a 95 percent confidence interval, our county-level estimates could also imply that a 100 percentage point reduction in the detention rate would decrease the prime working-age employment rate by 2.6 percentage points. Given this wide range of potential estimates, spillover effects may be present but cannot be definitively identified with the existing data.

We can similarly conduct a crude comparison of direct individual-level and aggregate estimates for Black individuals. Among Black respondents, the cumulative probability of arrest by age 28 is 40 percent (McCauley 2017), which we take as a rough benchmark for the size of the Black at-risk population. If we assumed a 100 percentage point reduction in the detention rate for this at-risk population, the estimates from Dobbie, Goldin, and Yang (2018) imply that the aggregate employment rate for Black individuals would correspondingly rise by 3.76 percentage points (9.4 percentage points times .40) through direct effects. However, our aggregate estimates of the association between changes in detention rates and changes in employment rates for Black individuals with baseline controls imply that a 100 percentage point reduction in the detention rate would increase the Black working-age employment rate by 40.8 percentage points (see panel B of table 2). Again, however, large standard errors mean that a large range

of estimates are possible. For example, within a 95 percent confidence interval, we cannot rule out that a 100 percentage point reduction in the detention rate would increase the Black working-age employment rate by 17.1 percentage points. This back-of-the-envelope calculation suggests potentially large spillover effects but we caution that these comparisons are highly speculative.

III.C. County-Level Detention Rates and Intergenerational Mobility

We conclude this section by considering the relationship between pretrial detention and intergenerational mobility among children, an important example of an intergenerational spillover effect. We estimate the intergenerational effects of pretrial detention by comparing county-level 1990 pretrial detention rate levels and measures of intergenerational mobility obtained from Opportunity Insights, as used by Chetty and others (2018). Based on SCPS data, in 1990, counties like Fulton, Georgia, and Orange, California, have the highest rates of pretrial detention with rates around 70 percent or higher, while counties like Suffolk, Massachusetts, and Essex, New Jersey, have the lowest rates of detention, with detention rates generally below 10 percent. We explore these relationships among thirty-nine counties in the SCPS for whom we can observe detention rates in 1990.[9]

Exploiting this cross-sectional variation, we present scatterplots and regression estimates of the following county-level specification:

$$IM_{c,2014} = \alpha + \beta_1 * Detention_{c,1990} + \beta_2 * X_{c,2000} + E_c$$

where $IM_{c,2014}$ represents the predicted mean percentile rank of income for children born between 1978 and 1983 to parents at the 25th percentile in the national household income distribution when they are age 31 to 37 (as measured in 2014 to 2015) in county c. $Detention_{c,1990}$ represents the 1990 detention rate in county c (in percentage points), and $X_{c,2000}$ represents baseline county-level covariates. We specifically choose to correlate mobility with the 1990 county pretrial detention rate as this captures the exposure that a child would experience at age 7 to 12. Any relationship between this pretrial detention rate and mobility likely reflects an intergenerational spillover effect given that these children are too young to be detained in adult pretrial systems. We caution that β_1 should generally not be interpreted as a causal estimate given that we are unable to control

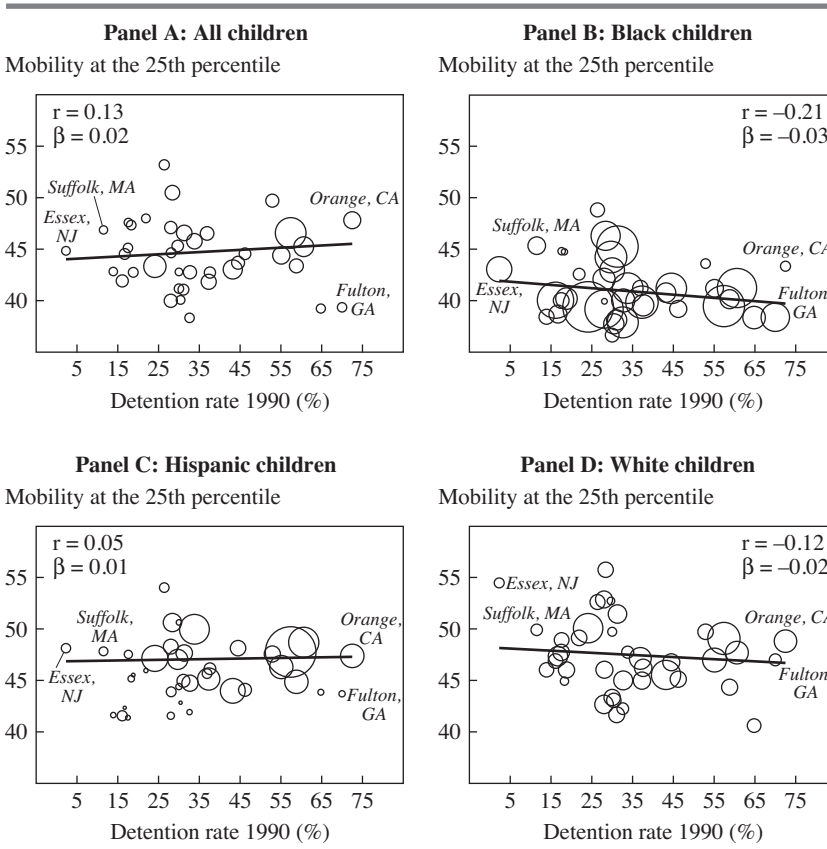
9. Summary statistics on 1990 detention rates and measures of intergenerational mobility are presented in panels A, D, and E of online appendix table A2.

for all potential differences between low- and high-detention counties, but we again feel these estimates are helpful suggestive evidence in the absence of convincing quasi-experimental variation in aggregate pretrial detention rates during this time period.

Figure 5 presents these findings and reveals a positive correlation of 0.13 between a county's 1990 rate of pretrial detention and intergenerational mobility for all children born to parents at the 25th income percentile and a linear regression coefficient of 0.02. However, this correlation becomes negative once one focuses in on Black children born to parents at the 25th income percentile. Among Black children, the correlation is −0.21 and the linear regression coefficient is −0.03. Counties with very high levels of pretrial detention rates in 1990, such as Fulton, Georgia, have a predicted mean percentile rank for Black children of 38.39. In contrast, counties with low levels of pretrial detention in 1990, such as Suffolk, Massachusetts, have a predicted mean percentile rank for Black children of 45.36. These regression estimates imply that for Black children, a one standard deviation increase in the 1990 detention rate (16.9 percentage points) is associated with a decrease in the predicted mean percentile rank of 0.54. For Hispanic children, a one standard deviation increase in the 1990 detention rate is associated with an increase in the predicted mean percentile rank of 0.1 and for non-Hispanic white children, a one standard deviation increase in the 1990 detention rate is associated with a decrease in the predicted mean percentile rank of 0.35. These results suggest that a characteristic of a high-mobility county may be its rate of pretrial detention, although we again note that our estimates are imprecise and noncausal in nature.

In panel C of table 2, we present estimates from a regression of mobility on 1990 detention rates with and without baseline controls. County baseline controls include the Gini coefficient for the bottom 99 percent, high school dropout rate, share black, share single mothers, social capital index, and violent and total crime rates, following Chetty and others (2014). We weight these regressions with the relevant county-level population for each racial and age group. After accounting for these baseline controls, we find that a one standard deviation increase in the 1990 detention rate (16.9 percentage points) is associated with a decrease in the predicted mean percentile rank of 1.12 for all children, a decrease in the predicted mean percentile rank of 0.93 for Black children, a decrease in the predicted mean percentile rank of 0.68 for Hispanic children, and a decrease in the predicted mean percentile rank of 1.89 for non-Hispanic white children. In figure 6 and panel D of table 2, we present analogous results measuring mobility for

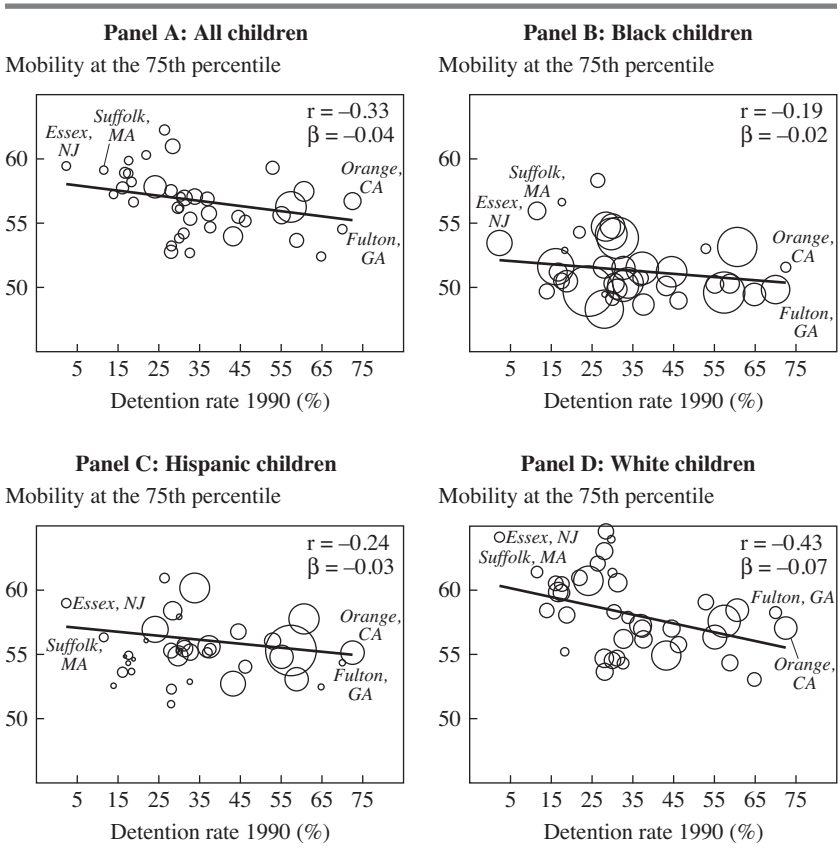
Figure 5. County Detention Rates and Mobility at the 25th Percentile for Parental Income



Sources: State Court Processing Statistics (SCPS); Opportunity Insights.

Note: The county detention rate is measured using the share of arrested felony defendants detained pretrial in 1990 in the SCPS for thirty-nine of the nation's seventy-five most populous counties. The income percentiles for children born 1978–1983 with parents at the 25th percentile are measured using the predicted mean percentile rank for children in the individual distribution of household income in 2014–2015 born to parents at the 25th percentile in the national household income distribution in each corresponding county. Correlations and best-fit regression lines are weighted using the applicable race-specific county population total as reported in the 2000 Decennial Census. The size of the circles represents the race-specific county population in 2000. See table 2 for additional regression estimates and standard errors.

Figure 6. County Detention Rates and Mobility at the 75th Percentile for Parental Income



Sources: State Court Processing Statistics (SCPS); Opportunity Insights.

Note: The county detention rate is measured using the share of arrested felony defendants detained pretrial in 1990 in the SCPS for thirty-nine of the nation's seventy-five most populous counties. The income percentiles for children born 1978–1983 with parents at the 75th percentile are measured using the predicted mean percentile rank for children in the individual distribution of household income in 2014–2015 born to parents at the 75th percentile in the national household income distribution in each corresponding county. Correlations and best-fit regression lines are weighted using the applicable race-specific county population total in 2000 as reported in the 2000 Decennial Census. The size of the circles represents the race-specific county population in 2000. See table 2 for additional regression estimates and standard errors.

children born to parents at the 75th income percentile, where we continue to find a general negative association between a county's 1990 pretrial detention level and mobility of children of all races.

POLICY IMPLICATIONS We can again use this type of cross-sectional evidence to simulate the two types of policy counterfactuals utilized previously. We first evaluate the impact on individuals of a change in detention rates, shifting rates in high-detention counties to match those in low-detention counties, before assessing what would happen if all counties reduced detention rates to only 10 percent via elimination of money bail relative to actuality.

For example, if Fulton, Georgia, reduced its detention rate to match the detention rate in Suffolk, Massachusetts, in 1990, there could be up to a 1.25 percentile decrease or a 3.84 percentile increase in the mean predicted percentile rank for all children born to parents at the 25th income percentile. Compared to Fulton's actual mean percentile rank of 39.35, its counterfactual percentile rank for all children would be 38.10 to 43.19. For Black children born to parents at the 25th income percentile, there would be an associated 1.84 to 3.20 percentile increase in the mean predicted percentile rank. Compared to Fulton's actual mean percentile rank of 38.39 for this demographic group, its counterfactual percentile rank for Black children would be 40.24 to 41.59.

Increases in intergenerational mobility would also occur if detention levels for all counties were reduced to 10 percent, as could be achieved by eliminating money bail. For intergenerational mobility, the mean predicted percentile rank of children across all counties in the SCPS is 44.82 for all children and 40.89 for Black children born to parents at the 25th income percentile. Applying our cross-sectional estimates, if these counties had reduced their 1990 detention rates to 10 percent, the counterfactual mean predicted percentile rank across all counties would have been 44.18 to 46.81 for all children and 41.84 to 42.54 for all Black children.

In addition, these simulations suggest that reforms like the elimination of money bail may also yield improvements in racial gaps in intergenerational mobility. For example, the mean predicted percentile rank is 47.43 for white children and 40.89 for Black children born to parents at the 25th percentile. Thus, the white-Black racial gap in mobility is 6.54 percentile ranks. As mentioned above, if counties eliminated money bail, our estimates suggest that the counterfactual mean predicted percentile rank for these same groups could instead be 41.84 to 42.54 for Black children and 48.05 to 50.81 for white children. Using the lower end of these counterfactual estimates, this policy reform, while benefiting both

racial groups, could also potentially reduce racial gaps in intergenerational mobility to 6.21 percentile ranks.

IV. Conclusions and Areas for Future Work

The US pretrial system has dramatically expanded over the past several decades and affects more than 10 million arrested individuals each year. The increasingly high rate of monetary bail coupled with the low financial resources of many arrested individuals has resulted in high rates of pretrial detention among these individuals, particularly for low-income minority populations. Much work remains to be done to understand the economic consequences of this pretrial system. While some recent research has started to measure the individual-level effects on individuals detained at the margin, rigorous work studying the potential spillover effects on families and community members is much needed.

This paper describes several pieces of evidence that can provide helpful guidance for policymakers. First, we document the significant direct consequences of pretrial detention on individual economic outcomes such as formal labor market attachment and the receipt of social benefits such as UI and the EITC. Second, we exploit county-level changes to show that these adverse consequences are also present in aggregate measures of economic well-being that incorporate spillover effects on other individuals. Finally, we provide more tentative evidence that pretrial detention may reduce the economic mobility of children. Put together, these three pieces of evidence indicate that reducing the scope of the pretrial system, such as through the elimination of money bail, is likely to generate significant economic returns for both directly affected individuals and the communities they live in.

ACKNOWLEDGMENTS We thank Adrien Auclert, Gabriel Chodorow-Reich, Janice Eberly, Supreet Kaur, Conrad Miller, Justin Wolfers, and numerous conference participants for helpful comments and suggestions. Jeremy Albright and Alexia Olaizola provided excellent research assistance.

References

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2012. "Do Judges Vary in Their Treatment of Race?" *Journal of Legal Studies* 41, no. 2: 347–83.

Agan, Amanda Y., and Sonja B. Starr. 2018. "Ban the Box, Criminal Records, and Statistical Discrimination: A Field Experiment." *Quarterly Journal of Economics* 133, no. 1: 191–235.

American Bar Association. 2007. *ABA Standards for Criminal Justice: Pretrial Release*, 3rd edition. Chicago: Author.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." *Quarterly Journal of Economics* 127, no. 2: 1017–55.

Arnold, David, Will Dobbie, and Peter Hull. 2020. "Measuring Racial Discrimination in Bail Decisions." Working Paper 26999. Cambridge, Mass.: National Bureau of Economic Research.

Arnold, David, Will Dobbie, and Crystal S. Yang. 2018. "Racial Bias in Bail Decisions." *Quarterly Journal of Economics* 133, no. 4: 1885–932.

Baughman, Shima Baradaran, and Frank McIntyre. 2012. "Predicting Violence" (February 6, 2011). *Texas Law Review* 90:497, 506–7.

Bayer, Patrick, and Kerwin Kofi Charles. 2018. "Divergent Paths: A New Perspective on Earnings Differences between Black and White Men since 1940." *Quarterly Journal of Economics* 133, no. 3: 1459–501.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2016. "The Masking of the Decline in Manufacturing Employment by the Housing Bubble." *Journal of Economic Perspectives* 30, no. 2: 179–200. https://www.aeaweb.org/articles?id=10.1257/jep.30.2.179.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2018. "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility." Working Paper 25147. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25147.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104, no. 9: 2633–79.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." Working Paper 19844. Cambridge, Mass.: National Bureau of Economic Research.

Demuth, Stephen. 2003. "Racial and Ethnic Differences in Pretrial Release Decisions and Outcomes: A Comparison of Hispanic, Black, and White Felony Arrests." *Criminology* 41, no. 3: 873–908.

Demuth, Stephen, and Darrell Steffensmeier. 2004. "The Impact of Gender and Race-Ethnicity in the Pretrial Release Process." *Social Problems* 51, no. 2: 222–42.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108, no. 2: 201–40.

Dobbie, Will, and Crystal S. Yang. 2019. "Proposals for Improving the U.S. Pretrial System." Washington: The Hamilton Project, Brookings Institution.

Doleac, Jennifer L., and Benjamin Hansen. 2020. "The Unintended Consequences of 'Ban the Box': Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden." *Journal of Labor Economics* 38, no. 2: 321–74.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie. 2019. "Judging Judge Fixed Effects." Working Paper 25528. Cambridge, Mass.: National Bureau of Economic Research.

Grogger, Jeffrey. 1995. "The Effect of Arrests on the Employment and Earnings of Young Men." *Quarterly Journal of Economics* 110, no. 1: 51–71.

Heaton, Paul, Sandra Mayson, and Megan Stevenson. 2017. "The Downstream Consequences of Misdemeanor Pretrial Detention." *Stanford Law Review* 69, no. 3: 711–94.

Heckman, James J., and Edward Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica* 73, no. 3: 669–738.

Holzer, Harry, Steven Raphael, and Michael Stoll. 2006. "Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers." *Journal of Law and Economics* 49, no. 2: 451–80.

Holzer, Harry, Steven Raphael, and Michael Stoll. 2007. "The Effect of an Applicant's Criminal History on Employer Hiring Decisions and Screening Practices: Evidence from Los Angeles." In *Barriers to Reentry? The Labor Market for Released Prisoners in Post-Industrial America*, edited by Shawn Bushway, Michael Stoll, and David Weiman. New York: Russell Sage Foundation.

Imbens, Guido, and Joshua Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62, no. 2: 467–75.

McCauley, Erin J. 2017. "The Cumulative Probability of Arrest by Age 28 Years in the United States by Disability Status, Race/Ethnicity, and Gender." *American Journal of Public Health* 107: 1977–81.

McConnell, Brendon, and Imran Rasul. 2018. "Racial and Ethnic Sentencing Differentials in the Federal Criminal Justice System." In *American Economic Association Papers and Proceedings*, vol. 108. Pittsburgh, Pa.: AEA Publications.

McIntyre, Frank, and Shima Baradaran Baughman. 2013. "Race, Prediction, and Pretrial Detention." *Journal of Empirical Legal Studies* 10, no. 4: 741–70.

Moffitt, Robert A. 2012. "The U.S. Employment-Population Reversal in the 2000s: Facts and Explanations." Working Paper 18520. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w18520.

Mogstad, Magne, Alexander Torgovitsky, and Christopher Walters. 2019. "Identification of Causal Effects with Multiple Instruments: Problems and Some Solutions." Working Paper 25691. Cambridge, Mass.: National Bureau of Economic Research.

Mueller-Smith, Michael. 2015. "The Criminal and Labor Market Impacts of Incarceration." Working Paper. https://sites.lsa.umich.edu/mgms/wp-content/uploads/sites/283/2015/09/incar.pdf.

Neal, Derek, and Armin Rick. 2014. "The Prison Boom and Lack of Black Progress after Smith and Welch." Working Paper 20283. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w20283.

New York City Criminal Justice Agency. 2014. *Annual Report 2013*. New York: Author.

Norris, Sam. 2019. "Examiner Inconsistency: Evidence from Refugee Appeals." Working Paper 2018-75. Chicago: Becker Friedman Institute for Economics, University of Chicago. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3267611.

Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108:937–75.

Pinto, Nick. 2015. "The Bail Trap." *New York Times*, August 13. https://www.nytimes.com/2015/08/16/magazine/the-bail-trap.html.

Raphael, Steven, and Sandra V. Rozo. 2019. "Racial Disparities in the Acquisition of Juvenile Arrest Records." *Journal of Labor Economics* 37, no. S1: S125–59.

Reaves, Brian A. 2013. "Felony Defendants in Large Urban Counties, 2009—Statistical Tables." Washington : US Department of Justice.

Rehavi, M. Marit, and Sonja B. Starr. 2014. "Racial Disparity in Federal Criminal Sentences." *Journal of Political Economy* 122, no. 6: 1320–54.

Rodgers, William M., III. 2019. "Race in the Labor Market: The Role of Equal Employment Opportunity and Other Policies." *Russell Sage Foundation Journal of the Social Sciences* 5, no. 5: 198–220.

Schlesinger, Traci. 2005. "Racial and Ethnic Disparity in Pretrial Criminal Processing." *Justice Quarterly* 22, no. 2: 170–92.

Shannon, Sarah K. S., Christopher Uggen, Jason Schnittker, Melissa Thompson, Sara Wakefield, and Michael Massoglia. 2017. "The Growth, Scope, and Spatial Distribution of People with Felony Records in the United States, 1948–2010." *Demography* 54, no. 5:1795–818.

Shierholz, Heidi. 2009. "At 10.2%, October's Unemployment Is a Wake-up Call." Economic Policy Institute, November 6. https://www.epi.org/publication/at_10-2_octobers_unemployment_is_a_wake-up_call/.

Smith, Sandra Susan. 2021. "The Difference a Day Makes: Considering How Pretrial Detention Informs Future Criminal Justice Involvement." Unpublished Manuscript.

Smith, Sandra Susan, and Nora C. R. Broege. 2020. "Searching for Work with a Criminal Record." *Social Problems* 67, no. 2: 208–32.

US Department of Justice. 2016. Brief for the United States as Amicus Curiae in *Maurice Walker v. City of Calhoun, Georgia*.

Western, Bruce. 2002. "The Impact of Incarceration on Wage Mobility and Inequality." *American Sociological Review* 67, no. 4: 526–46.

Comments and Discussion

COMMENT BY

CONRAD MILLER Will Dobbie and Crystal Yang characterize the economic costs of pretrial detention with a focus on its consequences for the labor market and economic insecurity. They discuss both the micro effects of pretrial detention on the detained and the macro effects on the broader community. At the micro level they summarize findings from their seminal paper, Dobbie, Goldin, and Yang (2018), where they find that pretrial detention increases the chances of conviction and decreases formal employment rates, earnings, and public benefits receipt using data from Philadelphia and Miami. At the macro level they present new findings on the county-level correlates of pretrial detention rates. They document that counties with larger increases in pretrial detention rates experience larger reductions in employment rates and increases in poverty rates.

The county-level results are particularly provocative. For example, the authors find that a 10 percentage point increase in county pretrial detention rates between 2000 and 2009 is associated with a 2 percentage point decrease in county employment rates for prime working-age adults.[1] The relationship is particularly strong and negative for Black employment rates. As Dobbie and Yang readily acknowledge, the aggregate relationship between pretrial detention rates and economic security is only suggestive of a causal relationship and likely reflects many confounding factors. One potential factor is that bail judges and magistrates set more generous bail conditions when defendants are gainfully employed. Another potential factor is that criminal justice jurisdictions with high pretrial detention rates may

1. County employment rates are calculated for 2000 and 2010, and exclude institutionalized individuals.

also be punitive along other dimensions, including conviction rates and confinement rates.

However, for the remainder of this discussion, I will assume that there is in fact a causal relationship between pretrial detention rates and employment rates and that the magnitude is economically important. My discussion will focus on what this causal relationship (should it exist) tells us about how employers consider applicants' criminal records in hiring decisions. Dobbie, Goldin, and Yang (2018) argue that pretrial detention lowers employment rates at least in part by increasing the chances that a defendant is convicted and that having a criminal conviction makes it harder to find a job. I will argue that the aggregate pattern suggests that employers are not primarily concerned about a job seeker's conviction because it signals something about worker productivity. If it were employers' primary concern, it is not clear why an exogenous increase in pretrial detention rates, and hence conviction rates, would significantly decrease aggregate employment. Instead, the aggregate pattern suggests that employers care about conviction status above and beyond its signaling content and that employers screen on conviction status directly, likely because convictions increase the perceived risk of negligent hiring lawsuits (Cavico, Mujtaba, and Muffler 2014; Lageson, Vuolo, and Uggen 2015) or restrict the set of tasks an employee can legally perform (Jacobs 2015).

Why does pretrial detention affect a defendant's labor market outcomes? In the short run, pretrial detention may lead to immediate job loss or disrupt educational attainment, housing, or family stability. As Dobbie, Goldin, and Yang (2018) argue, pretrial detention has longer-run effects on defendants' job prospects by increasing their chances of conviction, likely by weakening their bargaining position.[2] This conviction record follows the individual into the labor market.

These findings are consistent with an accumulating body of evidence that shows that having a prior conviction substantially harms an individual's job market prospects, regardless of the nature or length of the associated sentence (Pager 2003). Recent estimates indicate that about 25 percent of US adults have an arrest or conviction record (Jacobs 2015), and 13 percent of adult males have a felony conviction (Shannon and others 2017). We know that many employers use information on convictions in the hiring process; a recent survey of human resource professionals found that over 70 percent of firms conduct background checks for new hires (Holzer,

2. Heaton, Mayson, and Stevenson (2017) and Gupta, Hansman, and Frenchman (2016) also document that pretrial detention increases conviction rates.

Raphael, and Stoll 2004; Society for Human Resource Management 2018). While arrest records are more difficult to obtain in some states, court records on convictions are widely available (Bushway and Kalra 2021).

Many employers report that they are unwilling or reluctant to hire workers with criminal records (Holzer, Raphael, and Stoll 2004). A series of audit and correspondence studies provide compelling evidence that having a criminal record substantially reduces callback rates (Pager 2003; Agan and Starr 2018). This is true even for relatively minor offenses (Uggen and others 2014).

Mueller-Smith and Schnepel (2021) offer another compelling example. They study the criminal justice practice of diversion, where instead of getting a conviction on their record, defendants can avoid a conviction by successfully completing a probation sentence. In the setting they study, the marginal defendants who have their case diverted don't actually get a reduced sentence, they are just less likely to have a conviction on their record. Despite this seemingly artificial difference, the authors find large labor market gains to case diversion.

Most recently, Agan, Doleac, and Harvey (2021) study misdemeanor arrests and whether the prosecutor assigned to a defendant's case decides to file a court charge or declines to pursue the charge. Getting charged creates a criminal record and increases the chances of a conviction. But the authors study minor offenses where even defendants that get charged do not face significant punishment. They find that getting charged increases recidivism, as well as the chances that a defendant has a criminal record in the state repository. Again, a likely explanation for the increase in recidivism is that the presence of a criminal record worsens an individual's labor market prospects.

If we take for granted that pretrial detention affects a defendant's labor market outcomes primarily by increasing the chances of a criminal conviction, that leads to a natural follow-up question: Why do employers screen for a conviction in the hiring process? There are at least three reasons.

First, a natural labor economics view is that a conviction record can affect labor demand by serving as a negative signal of an individual's productivity (the productivity view). For example, employers may view applicants with criminal records as untrustworthy. Note that, for a prior conviction to be an informative signal, it must predict productivity above and beyond information on job applicants that is already readily available to employers.

A second view (the stigma view) is that a conviction record is associated with social stigma and that employers, employees, or customers would prefer not to employ, work with, or interact with individuals with convictions,

regardless of on-the-job productivity. This is akin to the "taste-based" discrimination in Becker (1957).

A third view (the legal costs view) is that employers prefer not to hire individuals with criminal records because of the legal restrictions or costs associated with a criminal conviction. Laws prevent those with certain convictions from working in some occupations. Employing workers with a prior conviction may increase an employer's vulnerability to a negligent hiring lawsuit. If an employee harms a coworker or customer, the employer may be held liable for damage if it can be shown that the employer was negligent in hiring that worker in the first place (Cavico, Mujtaba, and Muffler 2014). A prior conviction can be used as evidence for such negligence.

Survey evidence provides support for all three views (Society for Human Resource Management 2012; Lageson, Vuolo, and Uggen 2015). However, it is difficult to infer from employer survey responses alone what considerations drive hiring behavior (Pager and Quillian 2005). Determining which views are relevant is important because different mechanisms suggest different policy solutions. To the extent that employers are focused on the legal costs and restrictions of an employee's conviction status, policies that shape those costs and restrictions will influence labor demand for job seekers with conviction records.

I will argue that the notion that pretrial detention reduces employment rates in the aggregate is difficult to reconcile with the productivity view and most consistent with the legal costs view for why employers care about conviction records.

Suppose that local courts increase conviction rates while holding criminal conduct fixed, which is arguably the first-order effect of increasing pretrial detention rates. How would we expect this to affect the functioning of the labor market and how employers infer job seeker productivity in particular? We can view this policy as increasing the set of information about job seekers that is available to employers.[3] That's because conviction records are typically more readily available than arrest records (Jacobs 2015; Bushway and Kalra 2021). When conviction rates are low, there are more job seekers

3. Alternatively, we can view this policy as changing the categorization of job seekers rather than increasing the set of available information per se. For example, suppose the way the criminal court system works is that cases are ranked by the severity of the underlying criminal conduct and courts vary in the threshold they set for determining conviction. Suppose further that employers can observe who is convicted and some case details but not underlying criminal conduct. Then lowering the threshold for conviction increases the number of applicants for whom case details are available but may reduce the strength of the signal that a conviction provides.

with criminal records that some employers do not have access to. When conviction rates are high, those employers will have more information on arrests for the marginal convictions.

If we interpret an increase in conviction rates as increasing the set of information available to employers, it is not clear why this change would reduce employment rates in the aggregate. On the one hand, we would expect job seekers with those marginal convictions to have more trouble finding a job. On the other hand, when conviction rates are low, employers may try to infer the criminal history of job seekers from other available information, including applicants' job histories and personal characteristics. In the absence of objective information about who has a criminal record, employers will depend on their own subjective assessments of who is likely to have a criminal record. This behavior disadvantages job seekers who are stereotyped as likely to have a criminal record. Hence, an increase in conviction rates will help job seekers with no criminal history who may nonetheless be stereotyped as likely to have a criminal history when conviction rates are low.

There is an analogy here to the literature on ban-the-box policies, which prevent some employers from asking job applicants about their criminal history at the initial screening stage (Raphael 2021). Prominent research has argued that ban-the-box policies widen Black-white inequality in labor market outcomes by making it more difficult for employers to distinguish between Black applicants with and without criminal records (Agan and Starr 2018; Doleac and Hansen 2020). The argument goes that, in the absence of direct information on criminal history, employers may statistically discriminate against Black applicants. In the ban-the-box case, research indicates that a reduction in the criminal history information available to employers worsens Black labor market outcomes in particular. In this paper we have a change that arguably makes criminal history information more available, yet employment rates are reduced, particularly for Black adults.

Dobbie and Yang provide another clue that employers care about convictions per se rather than the information they convey about worker productivity. To identify the causal effect of pretrial detention Dobbie, Goldin, and Yang (2018) use what's known as a judge design—they take advantage of the fact that cases are essentially randomly assigned to bail judges or magistrates, and those judges and magistrates vary systematically in their tendency to detain defendants or subject them to monetary bail. They compare the outcomes of defendants who are assigned to low detention rate judges to otherwise similar defendants who are assigned to high detention rate judges. An important feature of this research design is that, for the marginal convictions the authors study, conviction actually conveys no

information about the defendant, including the defendant's productivity as a worker. Hence, if employers (a) only care about an applicant's conviction record to the extent that it conveys information about productivity and (b) can identify marginal convictions, then we may expect Dobbie, Goldin, and Yang (2018) to find no effect of pretrial detention on employment and earnings. The fact that they find large and negative effects suggests that either (a) or (b) does not hold.

In practice, employers cannot identify marginal convictions, at least not perfectly. (In fact, neither can the authors.) But when employers can access arrest records, prior work suggests that, conditional on initial charges, conviction is sufficiently arbitrary that it is not clear how much information it provides about worker productivity. The fact that labor market outcomes are so responsive to marginal convictions strongly suggests that employers care about convictions above and beyond their informational content.

While the findings Dobbie and Yang present are difficult to reconcile with the idea that employers use convictions in screening for their informational content, they are consistent with the view that convictions per se increase the legal costs of employing someone.[4]

Convictions restrict the set of occupations that individuals can legally work in (Jacobs 2015). For example, job seekers convicted of sex offenses may be banned from working with children. An increase in the conviction rates increases the set of job seekers subject to these legal restrictions.

A conviction on an employee's record makes employers more vulnerable to a negligent hiring lawsuit if that employee harms a fellow employee or a customer (Cavico, Mujtaba, and Muffler 2014). This increases the perceived risks associated with hiring a worker with a prior conviction. In both cases, it is easy to see why an increase in conviction rates would deteriorate labor market prospects for job seekers with criminal records and potentially decrease employment rates overall.

In summary, I interpret the evidence that Dobbie and Yang provide, in combination with prior work, as suggesting that employers care about a job seeker's conviction record per se above and beyond what it conveys about that job seeker's productivity on the job. If employers were primarily interested in the signaling content of a conviction record, it is not clear why an exogenous increase in pretrial detention rates, and hence conviction rates,

4. Under the stigma view, increasing pretrial detention increases the size of the stigmatized population. Whether we would expect this increase to reduce aggregate employment depends on whether there are sufficient nondiscriminatory employers (or employers serving nondiscriminatory customers) to absorb the increased number of stigmatized job seekers.

would meaningfully decrease aggregate employment. Instead, their findings support the view that employers are primarily concerned with the legal costs associated with a conviction record. This interpretation is subject to the caveat that the causal relationship between pretrial detention rates and aggregate employment may in fact be negligible or nonexistent. Dobbie and Yang make a convincing case that this aggregate relationship warrants further investigation.

REFERENCES FOR THE MILLER COMMENT

Agan, Amanda Y., Jennifer L. Doleac, and Anna Harvey. 2021. "Misdemeanor Prosecution." Working Paper 28600. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28600.

Agan, Amanda Y., and Sonja B. Starr. 2018. "Ban the Box, Criminal Records, and Statistical Discrimination: A Field Experiment." *Quarterly Journal of Economics* 133, no. 1: 191–235.

Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.

Bushway, Shawn D., and Nidhi Kalra. 2021. "A Policy Review of Employers' Open Access to Conviction Records." *Annual Review of Criminology* 4:165–89.

Cavico, Frank J., Bahaudin G. Mujtaba, and Stephen C. Muffler. 2014. "Criminal Background Checks in Employment: An Unfolding Legal Quandary for Employers." *Journal of Law and Criminal Justice* 2, no. 1: 41–103.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108, no. 2: 201–40.

Doleac, Jennifer L., and Benjamin Hansen. 2020. "The Unintended Consequences of 'Ban the Box': Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden." *Journal of Labor Economics* 38, no. 2: 321–74.

Gupta, Arpit, Christopher Hansman, and Ethan Frenchman. 2016. "The Heavy Costs of High Bail: Evidence from Judge Randomization." *Journal of Legal Studies* 45, no. 2: 471–505.

Heaton, Paul S., Sandra Gabriel Mayson, and Megan Stevenson. 2017. "The Downstream Consequences of Misdemeanor Pretrial Detention." *Stanford Law Review* 69, no. 3: 711.

Holzer, Harry J., Steven Raphael, and Michael A. Stoll. 2004. "Will Employers Hire Former Offenders? Employer Preferences, Background Checks, and Their Determinants." In *Imprisoning America: The Social Effects of Mass Incarceration*, edited by Mary Pattillo, David Weiman, and Bruce Western. New York: Russell Sage Foundation.

Jacobs, James B. 2015. *The Eternal Criminal Record*. Cambridge, Mass.: Harvard University Press.

Lageson, Sarah Esther, Mike Vuolo, and Christopher Uggen. 2015. "Legal Ambiguity in Managerial Assessments of Criminal Records." *Law and Social Inquiry* 40, no. 1: 175–204.

Mueller-Smith, Michael, and Kevin T. Schnepel. 2021. "Diversion in the Criminal Justice System." *Review of Economic Studies* 88, no. 2: 883–936.

Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108, no. 5: 937–75.

Pager, Devah, and Lincoln Quillian. 2005. "Walking the Talk? What Employers Say Versus What They Do." *American Sociological Review* 70, no. 3: 355–80.

Raphael, Steven. 2021. "The Intended and Unintended Consequences of Ban the Box." *Annual Review of Criminology* 4:191–207.

Shannon, Sarah K. S., Christopher Uggen, Jason Schnittker, Melissa Thompson, Sara Wakefield, and Michael Massoglia. 2017. "The Growth, Scope, and Spatial Distribution of People with Felony Records in the United States, 1948–2010." *Demography* 54, no. 5: 1795–818.

Society for Human Resource Management. 2012. *Background Checking—The Use of Criminal Background Checks in Hiring Decisions*. https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/pages/criminalbackgroundcheck.aspx.

Society for Human Resource Management. 2018. *Workers with Criminal Records: A Survey by the Society for Human Resource Management and the Charles Koch Institute*. https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/pages/second-chances.aspx.

Uggen, Christopher, Mike Vuolo, Sarah Lageson, Ebony Ruhland, and Hilary K. Whitham. 2014. "The Edge of Stigma: An Experimental Audit of the Effects of Low-Level Criminal Records on Employment." *Criminology* 52, no. 4: 627–54.

COMMENT BY

JUSTIN WOLFERS This paper by Will Dobbie and Crystal Yang is best understood as trying to draw policy implications from an important prior study they conducted with Jacob Goldin. That earlier study—Dobbie, Goldin, and Yang (2018)—found that detaining a defendant before their trial causes them to subsequently experience worse labor market outcomes. This conclusion follows from a clever quasi-experimental design which compared the long-run labor market outcomes of those defendants who were randomly assigned to lenient bail judges (who rarely require pretrial detention) with the outcomes of those defendants who were randomly assigned to less lenient judges (who were more likely to require pretrial detention).

The paper by Dobbie, Goldin, and Yang (2018) was a hit in the applied micro literature at least partly because of its methodological sophistication.

It was an early and convincing application of a "judge fixed effects" or "judge leniency" design, and it represents a cutting-edge application of the quasi-experimental methods that are often used in program evaluation. The present paper, in which the authors draw policy implications from that earlier analysis, can be considered a case study of the difficulty in drawing macro policy conclusions from even very well-identified micro econometric studies.

THE HAZARDS IN DRAWING POLICY RECOMMENDATIONS FROM WELL-IDENTIFIED CAUSAL ESTIMATES There are (at least) three sets of concerns that naturally arise when trying to draw policy conclusions from any empirical study. The first is the question of internal validity, which in this case means asking whether this research design yields a reliable estimate of the local average treatment effect for those marginal defendants affected by this natural experiment. This question tends to dominate debate within the academic applied micro research community, and indeed the "causality police" have been called to explore whether a judge fixed effects design really will yield internally valid inference about a local average treatment effect.[1]

Second is the question of external validity, which asks whether the findings from this natural experiment can be generalized to other settings and populations. Applied microeconomists have become much more interested in differences in treatment effects (treatment effect heterogeneity) in recent years. Dobbie and Yang are admirably clear that their earlier analysis in Dobbie, Goldin, and Yang (2018) only estimated the effect of pretrial detention on those "'marginal' defendants" (203) who might be released by a lenient judge but not by their stricter colleague. Yet the policy proposals they evaluate—such as the elimination of cash bail—are much more dramatic than these marginal changes, as they would almost eliminate pretrial detention. This is a problem because their prior study evaluated the consequences of allowing pretrial release of those defendants who at least some judges would have recommended be released, but in the present paper they

1. Frandsen, Lefgren, and Leslie (2019) provide a useful discussion of just how subtle the identification assumptions are. In brief, identification is straightforward if a strict judge would detain all those defendants that a lenient judge would, plus a few more. But if judges vary not only in leniency (that is, how many defendants they would detain) but also in who they judge important to detain, then the people who are detained by a strict judge are different from those detained by a lenient judge. (Formally, if the quasi-random judge assignments are instrumental variables, this is a violation of the usual monotonicity assumption.) In this case, the judge fixed effects design will yield estimates of the local average treatment effect that are confounded by differences in the average treatment effect across the groups of people that each judge would detain.

are extrapolating those findings to a population that includes defendants for whom all judges might otherwise have opposed pretrial release. It seems likely that this extrapolation is too optimistic about the effects of policy interventions that would largely eliminate pretrial detention.

Third, for a study to be relevant to policy, it must account for the full range of effects and not just the direct effects on the policy's direct beneficiaries. Dobbie and Yang note that their earlier study only identified the causal effect on an individual of that person receiving pretrial detention—which they call the direct effect—while there may also be unmeasured spillover effects to consider. As Baird and others (2014) note, "the impact of a program only on its beneficiaries becomes an unsatisfying answer to the real policy impact" (1). This problem of unmeasured spillover effects is sometimes also described as a problem of construct validity, as the measurements that were the focus of Dobbie, Goldin, and Yang (2018)—in this case only of the direct beneficiaries of a policy—are not the measurements that a policymaker would seek to rely on. None of this is news to Dobbie and Yang, and the present paper can be read as an attempt to address this set of concerns. The task they set themselves is to supplement their prior analysis with research that takes account of the relevant spillover effects.

But what are the most important spillovers in the present context? Dobbie and Yang highlight a set of socially mediated spillovers, arguing that "pretrial detention is likely to generate spillover effects . . . given its potential impact on families and communities." They argue that these socially mediated spillovers are likely negative, "as the costs of paying money bail and other related court fees and fines often fall on other family and community members of detained individuals." And so it follows that these "harms" amplify the direct effects. Thus, the authors could argue that their prior estimates of the direct effect of pretrial detention provide a lower bound for the aggregate effects. (Much of their text reads as if this is the implicit hypothesis.)

But as I'll argue below, there are also important market-mediated spillovers which likely attenuate the direct effect. Indeed, in standard models of the labor market, these market-mediated spillovers may even completely offset the direct effect. The idea is simply that if pretrial detention doesn't cause a shift in aggregate labor demand, then a job that a former detainee doesn't get is a job that goes to someone else. And so focusing on these market-mediated spillovers might lead one to argue that well-identified estimates of the direct effect are instead an upper bound on the aggregate effects.

When competing conjectures lead one to believe that Dobbie, Goldin, and Yang (2018) is either an upper bound or a lower bound of the aggregate effect of pretrial detention, it becomes clear that the prior study does not speak particularly clearly about the relevant policy issues.

It's worth pausing for a moment on this point, because this problem applies not just to Dobbie, Goldin, and Yang's (2018) careful study, but to literally hundreds of very careful and credible applied micro studies. After all, the vast majority of natural experiment papers focus on a causal estimate on individual people, families, businesses, or some relatively small subpopulation and so fail to account for spillover and general equilibrium effects. The difficulties in the present paper are simply a proxy for the broader question of how to get well-identified quasi-experimental methods to speak to policy questions. The current paper should be read as an attempt to solve this conundrum.

There are two ways forward. One strategy is to take seriously prior measures of the direct effect in Dobbie, Goldin, and Yang (2018) and then supplement them with further analysis that aims to directly measure, model, or otherwise bound these spillovers. This strategy builds on the existing evidence about direct effects and so is particularly appropriate when that prior evidence is of high quality. Chodorow-Reich (2019) is an example of this approach, as he describes conditions under which well-identified prior measures of the effect of state fiscal shocks on local state economies provide a lower bound for a particular national multiplier.

The alternative approach is to measure the aggregate effect of changing rates of pretrial detention, effectively estimating the combined consequences of both the direct and spillover effects. This is the strategy that Dobbie and Yang pursue, even though it means discarding their earlier work. The value of this strategy rests on a judgment about how convincing this new research is, relative to the prior but incomplete evidence. Given the high quality of Dobbie, Goldin, and Yang (2018), it might have been valuable to build on rather than replace that earlier work, even though doing so would have yielded an entirely different paper. The strength of the counterargument to this rests on the present paper exploiting a credible design to uncover convincing and statistically precise causal estimates.

RESEARCH DESIGN STILL MATTERS To make progress in measuring the aggregate effects, Dobbie and Yang (implicitly) assume that spillovers are limited to within a county. As a result, the total effect of a policy can be measured by comparing county-wide outcomes in those areas that are (quasi-)randomly assigned to a new treatment (such as eliminating cash bail)

and then comparing them to a control group with no such policy change. So far, so good.

This does not eliminate the need for a credible research design. Indeed, pushing the analysis to a higher level of aggregation simply shifts the research design challenge from the one set in Dobbie, Goldin, and Yang (2018) of finding individuals who were quasi-randomly assigned to pretrial detention to one of finding counties which were quasi-randomly assigned to a new treatment limiting the use of pretrial detention.

Unfortunately, the present paper does not deliver on this score. The central regressions correlate the change in various economic outcomes over the period 2000–2010 across twenty-four large counties, with the corresponding change in county-wide rates of pretrial detention among felony defendants. Some of the specifications include controls, but these control variables are all in levels, while the dependent variable is in changes. As such, there exists an enormous number of potentially confounding social, economic, political, legal, or crime-related variables whose changes are not controlled for, any of which may be responsible for the observed correlation between changes in economic conditions and changes in the probability of a felony arrest leading to pretrial detention.

Program evaluation research that purports to estimate causal effects tends to follow a pretty standard script in which the authors describe their design and why the particular variation that they have isolated might be considered exogenous. But this paper makes no such case.

The independent variable of interest is the change over a decade in the share of felony defendants in a county who are detained before their trial, and the paper offers no explanation about what might be driving this variation. Changes in rates of pretrial detention might be driven by exogenous changes in policy, but the authors provide no evidence of this. It's also possible that the variation in pretrial detention reflects judges responding to an array of broader social, cultural, economic, legal, or political forces. It's also possible that the detention rate might vary even if detention policies don't change, perhaps due to composition effects: the independent variable is the aggregate rate of pretrial detention, which might change if the mix of defendants accused of drug, property, and other crimes changed (and this compositional effect could change the aggregate even if judges didn't change how they treated defendants within any specific category). As such, anything that changes the mix of arrestees might be driving the variation in the independent variable, including changes in criminal opportunities, changes in the alternative labor market opportunities available to potential criminals, changes in the supply of potential victims, changes in policing policy, or changes in demographics.

The point is that there is a long list of factors that might drive variation in pretrial detention rates, and many of these factors likely also shape the outcome variables that Dobbie and Yang study, such as employment, poverty, or intergenerational mobility. To the extent that these factors are not controlled for in the analysis, they are omitted variables that potentially bias the empirical findings.

Moreover, while the paper reports estimates both with and without controls, it's important to note that while both the dependent variables (like the change in employment) and the independent variable of interest (the change in detention rates) are first difference or "change" variables, the control variables (like mean household income, the unemployment rate, the share of the population in various demographic and education groups—all measured at baseline) are included only as levels. Thus the regressions contain no controls for changes in economic, crime, or other factors over time.

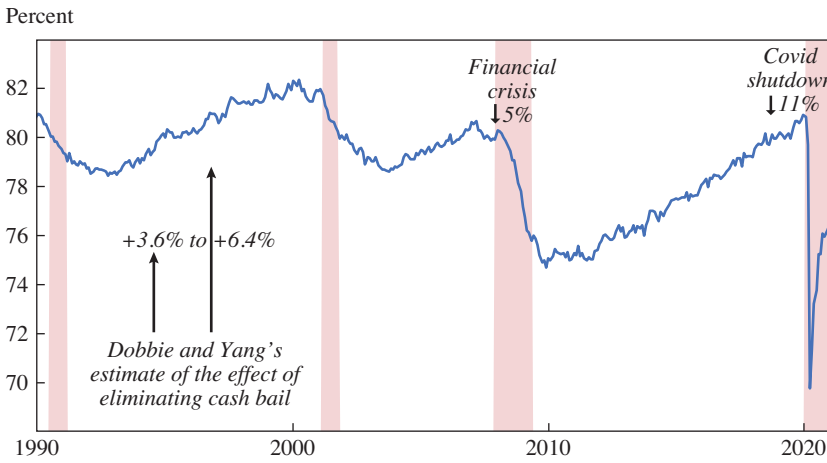
The paper offers the usual warning that the "analysis is exploratory in nature," to be interpreted with "an abundance of caution," and that "these county-level specifications should not be interpreted as precise or causal estimates" because it analyzes "county-level changes in detention rates that are likely endogenous," and so "one should be cautious in interpreting β_1 as a causal effect." These are warnings well worth heeding, although the authors appear not to do so, as the analysis then turns to the sort of counterfactual policy analysis that only makes sense if these new estimates are interpreted as a causal effect.

All of this presents a difficult trade-off for policy analysts. Is it better to base policy on well-identified estimates that omit any measure of spillovers or on estimates which do incorporate spillover effects but are likely biased? One might caution that basing policy recommendations on these new estimates rather than on Dobbie, Goldin, and Yang (2018) involves an unpleasant trade-off between some insight into spillovers and a study marred by omitted variable bias. It's easy to imagine that the resulting estimates are even less reliable for policymakers.

Indeed, as I demonstrate below, a careful assessment of the magnitudes suggests that the new estimates of the total effect of pretrial detention are implausibly large, uncomfortably imprecise, and yield a pattern across racial groups that seems quite improbable. This pattern suggests these estimates are influenced more by omitted variable bias than by the spillover effects the paper purports to estimate.

THE MAGNITUDES ARE INCREDIBLE To assess the magnitudes of the new Dobbie and Yang estimates, I will focus on their analysis of the effects of

Figure 1. Employment Rate of 25- to 44-Year-Olds



Source: Author's calculations using data from the Current Population Survey and the paper.

pretrial detention on the employment-to-population rate of 25- to 44-year-olds. I narrow my focus partly for brevity and partly because it is easier (at least for a labor economist) to interpret the magnitudes of employment changes. Many of the concerns that I raise apply in equal measure to Dobbie and Yang's analysis of changes in poverty rates and intergenerational mobility.

Table 2 of the paper shows the key result: regressing the change in the employment rate on the change in the detention rate yields a coefficient of −0.206 (without controls) or −0.115 (with controls). To interpret this magnitude, I'll focus on the key policy experiment that Dobbie and Yang analyze—the elimination of cash bail—which they argue would reduce pretrial detention rates from an average of 41.3 percent down to 10 percent. It follows that their regression predicts that this would raise the employment rate by $-0.206 \times (10\% - 41.3\%) = 6.4$ percentage points (and using the alternative coefficient from the regression with controls yields an effect of 3.7 percentage points). To be clear, these are effects on the employment rate in percentage points, and so relative to a typical prime-age employment-to-population ratio of around 80 percent, these estimates imply that employment levels would rise by 4.5 to 8 percent.

These numbers are implausibly large. To give some context, figure 1 shows the aggregate employment rate of 25- to 44-year-olds since 1990, highlighting periods of recession. The figure also superimposes arrows

showing the estimated effect of eliminating cash bail, so as to facilitate comparison with the size of business cycle fluctuations. The central estimates suggest that cash bail has a larger effect on aggregate employment than the 2001 "tech wreck" recession, and it is roughly comparable to the 2008 financial crisis, which was (at the time) thought to be a once-in-a-century shock.

While it is difficult to make direct comparisons of the employment effects of an incarceration effect to those from a financial crisis, the more recent pandemic-related shutdown provides a more directly relevant yardstick. The pandemic and associated lockdowns—which effectively forced millions of people to stay at home and shuttered the service sector of the economy—caused a sharp decline in the employment rate in which the prime-age employment rate fell by 11 percentage points between its peak in January 2020 and the deepest part of the trough in April 2020. Dobbie and Yang's estimates suggest that the institution of cash bail has an effect on employment that is of a similar order of magnitude, albeit about half as large. It is hard to believe that pretrial detention, which involves briefer lockdowns of a much smaller fraction of the population, could have employment effects that are of the same order of magnitude as an economy-wide shutdown.

Moreover, these calculations reflect the average result of this policy shift across the whole country. In those areas where pretrial detention rates are higher, eliminating cash bail would yield an even larger decline in detention rates, and so the authors' estimates effectively suggest that some states might enjoy much larger employment gains. For instance, online appendix table A2 suggests that the cross-county standard deviation of detention rates in 1990 was 17 percentage points, which is roughly half the average decline in detention rates if cash bail were eliminated. This implies that a county whose initial detention rate was one standard deviation higher than the average would be forecast to experience roughly a 50 percent larger effect than the average effects outlined above, and a county that was two standard deviations higher would have an effect that is twice as large as the average effect shown in figure 1.

THE IMPRECISION IS UNHELPFUL One response might be to counter that perhaps these counterfactual analyses involve taking the point estimates too seriously. Thus, rather than focusing on the point estimates, it might be worth focusing on the confidence intervals that surround them. Here, the regression without controls (which yielded a coefficient of −0.206) came with an estimated standard error of 0.109. Applying the resulting 95 percent confidence interval to the earlier extrapolation of the effects of eliminating

cash bail suggests that it would lead the employment rate to change somewhere between a –0.2 percentage point decline (which is only small relative to the numbers discussed earlier, but still a very large effect given the scale of pretrial incarceration) and an implausibly large 13.1 percentage point rise (in the specification without controls). The corresponding confidence interval from the specification including controls runs from a decline of –0.8 percentage points (which would be a large decline!) to an 8.0 percentage point rise.

Focusing on the confidence intervals yields a range that includes estimates that are no longer obviously implausible, but at the cost of suggesting that these estimates fail to falsify virtually any plausible effect.

The root cause of this statistical imprecision is not surprising: these regression results come from analyzing changes in only twenty-four counties over only a single time period. Moreover, the specification that includes controls adds eleven control variables, leaving very few degrees of freedom. A few power calculations at the beginning of the project might have led to the authors to look for an alternative empirical strategy.

ESTIMATED RACIAL DISPARITIES ARE TOO SMALL Dobbie and Yang probe beyond these aggregate effects and explore the differential effects of pretrial detention on employment (and poverty) rates by race and ethnicity. This analysis involves the same regressions as before, except the dependent variable is no longer the prime-age employment rate but rather the prime-age employment rate for a specific racial or ethnic group.

Importantly, the independent variable is the same in the regression analyzing Black employment as it is in the regression analyzing white employment: it is the county-wide detention rate averaged across all races, rather than a race-specific detention rate. This matters greatly for interpreting the coefficient estimates. If the social and economic processes that lead detention to affect employment are similar for Black people as for white people, then one might expect changes in the race-specific detention rate to have similar effects on Black and white employment. But changes in the aggregate detention rate—which is what Dobbie and Yang analyze—would then be expected to have quite disparate impacts on Black versus white populations, because Black people are dramatically overrepresented among detainees. After all, a policy that largely eliminates pretrial detention would lead a larger share of the Black population to avoid detention.

To get a sense of the relevant magnitudes, in the US population there are roughly five times more non-Hispanic white people than Black people, but among the population of detainees, there are 2.4 times more Black than non-Hispanic white people. Together, these numbers suggest that, on average,

a Black person is twelve times more likely to be detained than a white person. Thus, reducing or eliminating pretrial detention might be expected to have a much larger—perhaps twelve times larger—effect on the Black employment rate than the white employment rate.

Yet the coefficients that Dobbie and Yang report in table 2 suggest the effect of changes in the aggregate pretrial detention rate on the Black employment rate is only about one and a half to three times as large as the effect on the white employment rate (depending on which regression specification you prefer). This differential is surprisingly small, perhaps indicating the influence of other omitted variables.

To be a bit more precise, the idea here is simplest when thinking about the direct effect of detention on a defendant. Aggregating up to the level of a whole community, this direct effect would be expected to have an effect on the Black community that is twelve times larger, because Black defendants are twelve times more likely to be affected by a policy change. (Here, I'm following the authors in assuming the effects are linear.) The spillover effects are more complicated, because they might spread from a defendant of one race to a broader community that may be racially mixed. The more racially homogeneous one's community, the more likely it is that the spillover effects would also have a disproportionate effect by race. Consider first the social spillovers that concern Dobbie and Yang. If the Black and white communities never interacted, the social spillovers of pretrial detention would also be twelve times larger within the Black community, because a typical Black member of the community would be twelve times more likely to have a friend affected by changes in detention policy. If there were complete integration, then both Black and white defendants would be equally likely to have friends who were affected by changes in detention policy, and so there would be identical effects in the two communities. For the economic spillovers described below, the total number of jobs in the economy doesn't change and so one group's employment gains are another's losses. Thus, if the direct effect of eliminating pretrial detention were to raise Black employment rates twelve times more than white employment rates, and total employment is unchanged, then (assuming Black and white workers compete for the same jobs) Black and white workers would be roughly equally likely to have their employment prospects shaped by these spillover or equilibrium effects. These economic spillovers would lead the sum of the direct and indirect effects of eliminating or reducing pretrial detention to boost Black employment rates (largely through the direct effect) but decrease white employment rates (as white workers lose their jobs to Black workers).

The point here might be stated more simply as follows: pretrial detention has a radically disparate impact on Black and white communities, yet the results in the paper show only mild disparities. This suggests either problems with the estimates (like omitted variable bias) or that the results reflect a more subtle social process than that outlined by the authors.

A THEORY-INFORMED PRIOR I have argued that the empirical strategy pursued by Dobbie and Yang does not yield much insight into spillover effects, both because of the myriad ways in which the estimates are likely confounded by omitted variables and because of the imprecision of their estimates. While it is easy to harp on the problems with identification, the more important problem is how to constructively provide policy advice based on our limited knowledge. A careful causal study based on plausibly exogenous variation which accounted for spillover effects would be incredibly helpful. But in its absence, economists must still provide useful advice.

I would advise starting from a theory-informed prior about the likely sign and magnitude of the spillover effects, and this is where standard models of the labor market might be helpful. But first, a key fact to bear in mind is that pretrial detention has only a small effect on the number of days a defendant spends behind bars. Indeed, in Dobbie, Goldin, and Yang (2018), based on the judge leniency natural experiment, the authors estimated that an exogenously assigned period of pretrial detention leads a defendant to spend only an extra one to two weeks behind bars (see appendix table A12). Basically detainees are more likely to spend time behind bars before their trial—on average, an extra week or two—but at trial they're often granted sentences equal to time served, and so there is no effect on their post-disposition period of incarceration. With this fact in mind, the next question is what our standard labor market models predict would follow from changes in pretrial detention policy.

Competitive labor markets. The simplest approach might be to consider a competitive labor market. The labor demand curve is given by the marginal revenue product of labor, and the labor supply curve is dictated by the marginal utility of leisure; in equilibrium the wage adjusts to bring these into balance at a point where the quantity of labor demanded is equal to the quantity supplied.

In this framework, the presence or absence of pretrial detention will have no first-order effects on total employment, because both labor demand and labor supply are largely unchanged. The marginal product of workers is basically unaffected by the presence or absence of pretrial detention, and likewise the marginal utility of leisure is unaffected.

That's the first-order prediction with labor demand "basically unaffected" because there are no impacts on the marginal product of the vast majority of the workforce who are not arrested and hence not subject to detention. There is a fraction of detainees who will be affected, but even for them, the period of pretrial detention is so brief that it is unlikely to have noticeable effects on productivity. All told, the effect on labor demand is second order because there is only a small impact on the productivity of only a small fraction of the workforce. Similar reasoning suggests there are only second-order effects on labor supply. While labor supply is affected to the extent that there is an incapacitation effect—it's impossible to work while behind bars—on any given day the total share of the population that is undergoing a period of pretrial detention is tiny (while a significant proportion of the population are arrested at some point, because the period of pretrial detention is so short, few workers are detained on any given day). Moreover, many of those people at risk of pretrial detention are already largely detached from the workforce (according to Dobbie and Yang, "only 32 percent [of detainees] are employed in the year prior to arrest"). As such, a robust majority of the people who might be released were cash bail eliminated would not count as part of the labor supply under any detention regime.

Adding frictions. Of course, evaluating nonemployment in a perfectly competitive framework is somewhat limited given that model has no meaningful role for unemployment. A somewhat richer framework might allow for the sort of labor market frictions that create unemployment. One simple reduced-form approach is to posit that those frictions lead the real wage to get "stuck" above the level that would equate labor supply and labor demand. This simple formulation is a stand-in for a range of frictions, from minimum wage laws or other frictions that directly push the wage up, to union wage pressure or wage bargaining that creates a quasi-labor supply curve above labor supply, to efficiency wage concerns that create a quasi-labor demand curve above labor demand.

This simple framework yields the same stark insight: the presence or absence of pretrial detention has no first-order effects on total employment. Labor demand and labor supply don't shift (or barely shift) for the reasons articulated above. And pretrial detention does not directly shape any of the frictions laid out above (it won't affect the minimum wage, union wage demands, the no-shirking condition, etc.), and so it won't affect the real wage. As such, there's no effect on aggregate employment.

This prediction is still consistent with the key finding of Dobbie, Goldin, and Yang (2018) that pretrial detention reduced the detainee's future employment prospects. The reconciliation of no aggregate employment effect, even

with a substantial direct effect on detainees, is relatively straightforward, and it is all about the existence of an offsetting economic spillover. Once frictions create unemployment, they also create a metaphorical queue of potential workers at the factory gate looking for work. The factory owner now has more qualified applicants for each job than they need to hire, and so they can effectively discriminate against former detainees at no cost. After all, each detainee they don't hire can be replaced by hiring an equally talented non-detainee at the exact same wage. The simple point is that if pretrial detention policy doesn't change the number of jobs, then it won't have any effect on total employment, even as it affects who has those jobs.

This simple verbal model is akin to the ranking assumption of Blanchard and Diamond (1994), where employers who receive multiple acceptable applications might arbitrarily hire one set of candidates rather than another. In their setup, employers rank applicants by their unemployment duration; in the present case they might rank applicants by their criminal detention records instead. Blanchard and Diamond (1994) embed this assumption in a general equilibrium search and matching model that both incorporates matching frictions and dispenses with perfect competition in favor of Nash bargaining, and they still find that such "hiring rules do not affect how many are hired; but they determine who is hired, thus affecting the distribution of unemployment, as well as wages" (421).

CONCLUSION Dobbie and Yang have made a convincing case that pretrial detention is an important institution worthy of further study. Their earlier analysis pushed the issue to the forefront, showing quite large negative consequences at the individual level. The present paper seeks to go a step further, incorporating analysis of possible spillover effects.

Their new estimates suggest very large negative spillovers that yield extraordinarily large macroeconomic effects. But I'm not convinced, as I find the identification strategy weak and undefended, the racial pattern at odds with what one might expect, and the estimates imprecise.

The question a policy analyst is left with is whether to rely on these new estimates or use Dobbie, Goldin, and Yang (2018) as a more reliable starting point when analyzing the economic consequences of pretrial detentions. My sense is that the low statistical power in the present study combined with the risk that these new estimates are severely confounded by omitted variables means that point estimates could be quite some distance from the truth. Instead, I would suggest starting with the more credible estimates in Dobbie, Goldin, and Yang (2018), understanding that they need to be supplemented with some insight into likely spillovers. On this score, standard models of the labor market suggest that spillover effects likely operate

to somewhat attenuate the direct effects. This perspective suggests that Dobbie, Goldin, and Yang (2018) had identified a plausible upper bound on the total effects of pretrial detention, with zero as the corresponding lower bound. This yields a set of bounds on the total effects of pretrial detention that is smaller, narrower, and more plausible than in the present paper.

REFERENCES FOR THE WOLFERS COMMENT

Baird, Sarah, Aislinn Bohren, Craig McIntosh, and Berk Ozler. 2014. "Designing Experiments to Measure Spillover Effects." Working Paper 6824. Washington: World Bank. https://documents1.worldbank.org/curated/en/266671468089953232/pdf/WPS6824.pdf.

Blanchard, Olivier Jean, and Peter Diamond. 1994. "Ranking, Unemployment Duration, and Wages." *Review of Economic Studies* 61, no. 3: 417–34.

Chodorow-Reich, Gabriel. 2019. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *American Economic Journal: Policy* 11, no. 2: 1–34.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108, no. 2: 201–40.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie. 2019. "Judging Judge Fixed Effects." Working Paper 25528. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25538.

GENERAL DISCUSSION Will Dobbie responded to comments from Justin Wolfers on the data, clarifying that it was case-weighted and incorporated base changes. Dobbie said that, with this considered, Wolfers and the authors had the same numbers for their calculations. Dobbie agreed completely with Wolfers that the changes-on-changes identification strategy should not be interpreted with the same confidence as the micro estimates from Dobbie, Goldin, and Yang.[1]

Dobbie contested Wolfers's comparison of the results to those of Agan and Starr as not perfect because in that study, the characteristics of the people remained constant, while in this paper, the people are able to potentially avoid a scarring activity that changes their characteristics.[2]

1. Will Dobbie, Jacob Goldin, and Crystal S. Yang, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review* 108, no. 2 (2018): 201–40.

2. Amanda Agan and Sonja Starr, "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment," *Quarterly Journal of Economics* 133, no. 1(2018): 191–235, https://doi.org/10.1093/qje/qjx028.

Janice Eberly asked about the comparison between eliminating cash bail and banning the box. She further wondered if there might be other policies that would work toward alleviating the effects of pretrial detention.

Dobbie pointed out two ways to aid disadvantaged groups that are most affected by these policies. First, changing the cash bail system is an example of a policy that prevents scarring. Dobbie, Goldin, and Yang find that the impact on safety from lenient judges is minimal, while the economic impacts for the individual are large. Dobbie argued that the elimination of cash bail is an obvious policy to undertake because it makes these individuals more employable with minimal societal downside.

Dobbie said that the other set of policies that would be useful to alleviate the employment effects of a criminal record are policies that attempt to support reintegration into the economy despite scarring. Dobbie cites ban the box as one policy that fits into this category. Dobbie noted that the way that employers use signals like criminal record in the evaluation of employment applications is suboptimal, underscoring the importance of policies that improve the reintegration process. Dobbie also mentioned two other policies suggested by Conrad Miller: increased liability protection for employers and increasing the amount of objective information provided in the hiring process.

Dobbie has other ongoing work showing that workers with and without a criminal record are equally productive. Employers were receptive to being told about their mistake in avoiding workers who were equally productive.

Wolfers responded that he did not think ban the box was a perfect analogy for Dobbie's pretrial policy, but instead wanted to consider the general equilibrium with a straightforward model. Since the time that people spend in jail before going to trial is about two weeks, Wolfers thinks that the effect must be mostly from signaling. Wolfers gave the example of putting the letter *L* on the foreheads of 10 percent of the population and measuring the labor market effects. Clearly, the group that is discriminated against will face negative consequences, but Wolfers said that his strong prior, based on most models of the labor market, is that the overall employment effect must be close to zero if the effect comes from signaling. Wolfers acknowledged that there could be second-order effects that create an overall employment effect and emphasized the challenges in applying micro analysis that is very well identified to understanding what will happen in equilibrium.

Dobbie said that he would hesitate to accept that the employment effects are zero because there was also skepticism about the micro estimates in the work of Dobbie, Goldin, and Yang, showing that even though pretrial detention averages two weeks, labor market effects are still present after

four years. Dobbie agreed that this paper does not completely bridge the gap between the micro and macro effects of pretrial detention but also expressed optimism that it will be possible to make credible macro estimates in the future.

Miller commented that comparing pretrial detention policy to ban the box is useful for interpreting how the labor market is responding to criminal history. Miller noted that there is something contradictory about the way that employers respond to a criminal record as shown in ban the box but do not dig into the arrest record, which would allow a deeper understanding of the context for a conviction or lack of conviction.

Steven Davis brought up two issues. First, there is the question of why employers respond so strongly to a conviction. Davis said that Wolfers made an important point on this issue, that employers often cannot discriminate against convicted employees in the form of lower wages, which might make hiring them more desirable. Davis noted minimum wage, collective bargaining, and the threat of lawsuits as impediments to this potential employer incentive for taking a chance on new hires with convictions. Second, Davis agreed with Wolfers's point that the micro effects of detention are likely to be diluted in equilibrium; however, Davis suggested that there are likely to be longer-term effects on future human capital accumulation for workers who are scarred by pretrial detention. Due to this scarring, they will be more likely to have future convictions rather than pursuing human capital gains through education and on-the-job training. This effect can cumulate over time for the individual and persist in equilibrium, because a segment of the population becomes permanently less productive and thus less employable.

Erica Groshen first raised the finding from her Cornell colleagues' work that the information that employers have on criminal records is often inaccurate.[3] Accounting for the deficiencies in these records could make the overall labor market impacts worse. Groshen also said that considering displaced workers is another way to think about the impact.[4] If workers faced with pretrial detention lose their jobs and face scarring, it could have an impact that mirrors that of displaced workers, particularly those

3. Martin Wells, Erin York Cornwell, Linda Barrington, Esta Bigler, Hassan Enayati, and Lars Vilhuber, *Criminal Record Inaccuracies and the Impact of a Record Education Intervention on Employment-Related Outcomes*, Cornell Criminal Records Panel Study 2020-01; https://ecommons.cornell.edu/handle/1813/103780.

4. See, for example, Henry S. Faber, "Employment, Hours, and Earnings Consequences of Job Loss: U.S. Evidence from the Displaced Workers Survey," *Journal of Labor Economics* 35, no. S1 (July 2017): S235–S272.

with low levels of education. Groshen suggested that this would imply larger declines in wages and potentially lead to estimates of employment effects that were more in line with those from the paper than what Wolfers would predict.

Caroline Hoxby recounted Wolfers's description of the simple process by which one worker moves to unemployment and is replaced by a worker who was previously unemployed, commenting that under ban the box legislation, Agan and Starr found that there were increases in discrimination against Black workers. Hoxby noted that considering the racial distribution of employed and unemployed workers could lead to exaggerated spillover effects relative to ignoring the racial composition of those groups.

CAROLINE M. HOXBY
Stanford University

Advanced Cognitive Skill Deserts in the United States: Their Likely Causes and Implications

ABSTRACT I use mapping and age trajectories of advanced cognitive skills to better understand why these skills are more prevalent in some local areas than in others. The study begins by explaining what advanced cognitive skills are. It offers a nonspecialist's review of recent brain science that indicates that adolescence is the key period for the development of advanced cognitive skills. The paper considers three main explanations for why the prevalence of advanced cognitive skills varies substantially across US counties. Is it early childhood factors which could generate endogenous responses that are important later when advanced cognitive skills are developing? Is it factors whose influence is greatest during adolescence—the period when brain science argues that experience would most directly affect advanced cognitive skills? If so, adolescence is indeed the age of opportunity but also risk. Is the variation among counties explained by migration of individuals toward areas where other people have advanced cognitive skills similar to their own? Evidence based on cognitive skill trajectories, maps at different ages, and longitudinal regressions suggests that all three of these explanations play a role in generating areas where advanced cognitive skills are prevalent and areas where they are not—advanced cognitive skill deserts.

"The abilities that develop in adolescence . . . are not as necessary for survival as are those that develop early in life. You can live without being able to reason logically, plan ahead, or control your emotions (the plenitude of illogical, impetuous, and short-tempered adults attests to this). . . .

Conflict of Interest Disclosure: The author did not receive financial support from any firm or person for this paper or from any firm or person with a financial or political interest in this paper. She is currently not an officer, director, or board member of any organization with an interest in this paper.

"Unlike elementary skills, whose development is tightly regulated by preprogrammed biology, evolution left more room for variation in the development of complex abilities. That's why there's so much variation in how well different people reason, plan for the future, and control their emotions, but far less variation in how well people see, hear, and walk.

"In the past, not all environments demanded . . . advanced cognitive abilities. . . . In today's world, though, where formal education is increasingly important for success, people who are bad at reasoning, planning, and self-regulation are at a serious disadvantage, and the fact that the development of these abilities is highly sensitive to environmental influence is a mixed blessing. . . . For people . . . in favorable circumstances [during early adolescence], the plasticity of these brain systems is wonderful. For those who [aren't in such circumstances], this same plasticity can be disastrous" (Steinberg 2015, 29–30).

This quotation hits on three points. Each of them is important to this project. The first point is that numerous vital skills result from very early brain development. Steinberg mentions seeing, hearing, and walking. Neuroscience indicates that we might add language, social, and numerous other noncognitive skills to this early mix. Very early brain development mainly affects the back and center lobes of the brain and is, as Steinberg suggests, focused on skills that are crucial for survival and integration into a society. ("Very early" refers to the period starting approximately with the final three months of gestation and ending when a child is about age 3. Neuroscientific evidence based on brain scans is discussed in the next section.)

It is important to note that nowhere in this paper do I argue that the aforementioned skills are anything less than highly consequential for a person's life outcomes. Indeed, Steinberg (2015) is arguing that it is precisely because they are crucial that they differ relatively little among people, presumably owing to natural selection among humans over hundreds of thousands of years. Hereafter, I refer to the aforementioned skills as "noncognitive" as a shorthand common among economists, but some of these skills (such as early language development) clearly have strong cognitive elements, as well as social, visual, and hearing elements.

Second, the quotation states that advanced cognitive skills develop mainly during adolescence, when the frontal lobe of the brain is in its most intense period of transformation. Steinberg (2015) and others argue that the fact that advanced cognitive skills develop in adolescence cuts both ways.[1] On the

1. See, for instance, Reyna and others (2012) and Dahl (2004).

plus side, adolescence is a period when society has opportunities to intervene to improve a person's advanced cognitive skills. Also on the plus side is the fact that adolescence is a stage at which children have at least some agency. To put it plainly, adolescents spend a lot of time in schools that can govern their experiences to a considerable extent. However, adolescents can take at least some control of their lives so that they can possibly acquire advanced cognitive skills even if those around them do not support their endeavors. Family assignment luck thus potentially plays a smaller role in frontal lobe development than in very early brain development. This is plausibly good news for adolescents: a child who was initially unlucky could perhaps offset some lost noncognitive skills by acquiring advanced cognitive skills.

On the minus side, if the development of advanced cognitive skills is "highly sensitive to environmental influence," as Steinberg (2015, 30) argues, then the plasticity of the frontal lobe in adolescence makes this period a risky one.

Third, the quotation suggests that there are some environments in which advanced cognitive skills are necessary for economic success. In highly developed economies such as America's, an economist's mind naturally thinks of skill-biased technological change which, by definition, favors those with advanced skills. Globalization and immigration of the sorts that are specific to the United States also come to mind, though the degree to which these are skill-biased is complex and the subject of debate. In addition, one might think that some areas of the United States might be significantly more skill-biased than others if there are agglomeration economies associated with advanced cognitive skills. The economics literature contains many examples of agglomeration economies associated with noncognitive skills, such as physical endurance needed for certain manufacturing. Nevertheless, economists have suggested that there are certain economies where people with advanced cognitive skills will thrive more than people who lack them and there are models that would justify such agglomeration.[2] In addition, if people with advanced cognitive skills share preferences for certain amenities, that could amplify or even fully justify agglomeration.[3]

If we accept the idea that there are possibly agglomeration economies associated with advanced cognitive skills, then several phenomena are

2. For a recent review article on agglomeration economies, see Mori (2017). See also Kolko (2007) and Forman (2013) for agglomeration of high-skilled workers.

3. On this point, see Couture and others (2020)

potentially relevant. Skilled-biased technological change (and plausibly trade and immigration) could exacerbate the differences in outcomes between local economies where advanced cognitive skills are and are not prevalent. High school graduates with advanced cognitive skills may be especially likely to migrate to a residential college, and people with advanced cognitive skills are—almost by definition—prevalent in college towns. Adult migration may allow people to find an economy that suits their skills or tastes for amenities. There is some evidence that more educated people make relocation decisions that are more beneficial to them economically.[4] If children can only learn advanced cognitive skills from adults in their environment who have such skills themselves, then agglomeration economies may produce places that are substantially more or less propitious for advanced cognitive skills development. Of course, this mechanism might wane with the advance of modern communications technology which could, for instance, allow children to take classes remotely.

Having set out some key arguments and logic, I now turn to what this paper tries to do. First, it attempts to test whether there is much variation among US counties in the prevalence of adults with advanced cognitive skills.[5] Being interested in areas where advanced cognitive skills are especially non-prevalent, I label such areas "deserts" as a shorthand.

One might wonder why cognitive skills are important and why I do not pursue a far simpler task such as mapping educational attainment or the prevalence of jobs in high prestige occupations. After years of researching education and economics, I am convinced that it is skills that are foundational. Every model of human capital is a model of skills. Educational attainment is merely a proxy and can be a crude one because the mapping between academic degrees and skills varies widely among schools. One high school's diploma standards may be much higher than another's. Colleges vary even more in their degree standards, due in part to differences in admissions selectivity and consequent expectations for students. If we fail to focus on skills, we can fall into the trap of promoting policies that merely produce degrees. We can also struggle to explain why some people or areas, apparently with the same educational attainment, end up with very different employment, social, and political outcomes. We can then end up focusing on explanations that are actually second order. In short, I am interested in variation that is foundational: cognition.

4. See Malamud and Wozniak (2012).

5. Maps based on Commuting Zones look broadly similar but exacerbate some problems that already plague county-based maps.

Second, this paper attempts to provide enough age-related evidence on the three explanations to help us parse them. Specifically, are advanced cognitive skill deserts associated with early-age factors that can be difficult to affect with policy? Are they associated with adolescent-age factors more likely to be influenced by policy? Finally, are they associated with individuals migrating toward areas where other people have advanced cognitive skills similar to their own?

In the previous paragraph, the choice of the word "associated" was deliberate. This is not a paper that attempts to test whether specific policies, such as school resources, *cause* variation in the prevalence of advanced cognitive skills. I take up such causal questions in two related papers.[6] Even without attempting to show causality, I believe that the descriptive evidence in this paper is revealing.

The last section of the paper discusses some implications of the evidence described, including what we learn about the mechanisms most likely to improve advanced cognitive skills. I also provide a bit of speculative evidence on possible links between advanced cognitive skills, economic fatalism, social trust, and politics.

Finally, despite the use of the word "desert," this paper is written in the spirit of Steinberg's (2015) title, *Age of Opportunity*, with an emphasis on opportunity. One of the most fundamental problems in economics is how to maximize social welfare, given the constraints imposed by individuals' endowments. Of course, the problem is more complicated but it is differences in endowments that drive inequality in utility and set up many of the tensions between equity and efficiency. Skill-biased technical change, especially if it favors advanced cognitive skills, can exacerbate these tensions. Understanding why the development of advanced cognitive skills varies so much from one geographical area to another, as I will show in this paper, may be a first step toward causal means to improving such skills for potentially many people, thereby relaxing the constraints associated with endowments. The paper thus attempts to identify channels whereby societies could avoid the seeming inevitability of increasing strain between adults who do and do not have advanced cognitive skills.

6. These two other papers make up my Tanner Lectures on Human Values for the University of California, Berkeley. These lectures were postponed due to the coronavirus pandemic so, although they precede this paper as a matter of logic, they will actually be released after it. Briefly, the first paper argues that early adolescence is a "fork in the road" because students who begin acquiring advanced cognitive skills tend to stay on that trajectory and vice versa. The second paper uses several natural experiments to demonstrate that successful learning-related interventions are especially productive in early adolescence.

I. Advanced Cognitive Skills

I.A. What Are Advanced Cognitive Skills?

Advanced cognitive skills are generally defined as those that require higher order reasoning. They require a capacity to solve problems through logic, think in the abstract, engage in critical thinking, and derive general principles from a set of facts. They are also often described as integrative or synthesizing. Advanced cognitive skills can be meaningfully differentiated from skills such as memorization, summarization, organization of facts, and other methods of acquiring concrete information. These distinctions were recognized many years ago, and they have now attracted a vast body of research.[7]

For the purposes of economics, examples may be more helpful than a summary of the research. It is instructive to start with some mathematics examples because mathematics curricula are often associated with specific school grades. Such associations are helpful for developing intuition about the relationships between the age at which a skill is typically learned and the likelihood that the skill is an advanced cognitive one.

Long division, long multiplication, addition of fractions: these are all exercises that require a person to follow an algorithm that can even be fairly complex. They differ from algebra because algebra requires a person to translate a problem into equations, which are then solved to find the solution. Algebra thus requires abstract thinking. Proof-based mathematics is even more distinct because it requires logical reasoning that can extend over many steps. In short, algebra and the mathematics that typically follow it in the curriculum can be characterized as advanced cognitive skills.

Now consider coursework in history. Earlier coursework may require a student to condense facts, organize them around narratives, and recognize similarities and differences among historical events and personalities. However, a history course involves advanced cognitive skills if it requires students to analyze cause and effect by engaging in critical thinking and drawing abstract generalizations from facts. Also, a history course might involve advanced cognitive skills in writing and reading if it requires students to integrate and synthesize material.

7. Piaget (1972) is seminal; Chapman, Gamino, and Mudar (2012) provide an excellent review.

In this paper, I rely heavily though not exclusively on mathematics-based examples and measures. This is for the reason given above: mathematics curricula tend to follow a conventional sequence that makes it easier to follow as steps in cognitive development. Also, there are idiosyncratic data constraints, described below, that sometimes force me to rely on mathematics data. However, advanced cognitive skills are by no means restricted to mathematical reasoning and its near relatives. Indeed, verbal skills data tend to show similar results when they are available (see table 2). If there were measures of planning and strategic thinking available, I would wish to include them as well. One should think of mathematics as a good marker of cognitive skills development. It is by no means the only type of advanced cognitive skill.

I.B. Adolescent Brain Development in Brief

Advanced cognitive skills are associated with frontal lobe brain development, which is particularly dramatic starting with early adolescence. During this period, the frontal lobe (which has previously undergone exuberant growth of neural circuits) intensely prunes synapses, much as one might prune a tree to keep the strongest branches so that they can grow better. Pruning is followed by myelination, which is the development of specialized membrane around axons. Myelination essentially speeds up circuits that remain after the pruning. These stages may be thought of as the laying out of cognitive possibilities (exuberant development), the training of the brain about which possibilities need to be prioritized (pruning), and making the brain work faster on the prioritized possibilities (myelination). The brain is at its most plastic, most receptive to experience, when these stages take place most rapidly and dramatically.

These same stages begin before birth within the caudal brain stem (the brain stem being the posterior part of the brain). The most rapid and intense period of back and central brain myelination has already occurred by the end of the first two years of postnatal life. That is, key stages of brain development, such as pruning and myelination, occur both in very early childhood and adolescence, but the period in which the frontal lobe is *most* plastic begins much later, with most researchers suggesting 10 to 11 years of age for females and 11 to 12 for males.

Because the frontal lobe has long been associated with higher order reasoning, researchers have argued that the age at which training for advanced cognitive skills should optimally begin is the period in which this part of the brain is most plastic. Recent studies directly link adolescent anatomical

and functional brain changes to cognitive experiences and training. They depend especially on longitudinal use of magnetic resonance imaging since that noninvasive technique allows researchers to follow the same person's brain anatomy as he or she undergoes new learning experiences.[8] It should be mentioned, though, that much of this linking work is not convincingly causal because adolescents self-select into experiences that take place over an extended period—for instance, choosing which courses to take or whether to participate in chess club. Randomized controlled trials in which researchers manipulate adolescents' experiences naturally must occur over short time periods and are thus less revealing about longitudinal brain changes. I argue in a related paper for the use of natural experiments which allow for credible causal inference combined with a longer longitudinal study period.[9]

I.C. Advanced Cognitive Skills Trajectories in Early Adolescence

Psychiatry researchers and educators have long recognized that advanced cognitive skills can rapidly expand during adolescence, though they continue to be honed and to mature throughout adulthood. Nevertheless, not all adolescents experience this rapid expansion. Instead, cognitive neuroscientists and educators have observed that some students "stagnate cognitively and fail to thrive academically" in early adolescence: "With regard to vulnerability, the years when adolescents are in middle school (fifth through ninth grades) represent a period metaphorically referred to as a transitional 'black hole' in education" (Chapman, Gamino, and Mudar 2012, 124).

Cognitive neuroscientists tend to attribute such stagnation to frontal lobe development that is insufficiently rapid or dramatic. This, in turn, could be ascribed to learning experiences that are insufficiently rich or demanding to encourage frontal lobe development.

While frontal lobe development continues well past early adolescence, advanced cognitive skill trajectories are often somewhat set, in practice, by age 14 to 15 for females and age 15 to 16 for males. Although not from the aforementioned natural experiments, some evidence that early adolescence is crucial in setting these trajectories can be derived from longitudinal education data. In what follows, I show transition matrices based on two large longitudinal studies produced by the National Center for Education Statistics (NCES): the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), which covers children in kindergarten through

8. See Spear (2013) and Atkins and others (2012) for excellent reviews.
9. I am referring to one of the Tanner Lecture papers mentioned above.

Figure 1. Male Math Score Transitions, Ages 10.5 to 13.5

Deciles of math in grade 5 (age 10.5 on average)	1	2	3	4	5	6	7	8	9	10
1	67.4	23.1	5.4	1.9	1.4	0.5	0.3	0.0	0.0	0.0
2	29.2	32.9	22.2	9.3	3.5	1.5	0.3	0.6	0.6	0.0
3	10.6	23.6	25.8	18.7	11.7	6.5	2.4	0.3	0.5	0.0
4	3.6	12.7	19.0	21.6	20.0	11.4	7.0	3.9	0.8	0.0
5	1.5	3.4	11.1	17.6	24.1	21.7	11.1	6.8	2.4	0.5
6	1.0	1.4	8.4	11.2	20.1	18.4	19.1	11.0	8.1	1.2
7	0.0	1.0	4.0	7.3	13.0	16.5	21.6	19.3	12.0	5.3
8	0.0	1.4	0.9	3.9	6.1	15.0	19.1	22.7	20.2	10.9
9	0.2	0.2	0.8	1.0	3.3	6.6	14.9	23.6	26.9	22.4
10	0.0	0.0	0.0	0.2	0.8	2.2	6.2	13.6	24.6	52.4

Deciles of math in grade 8 (age 13.5 on average)

Source: Early Childhood Longitudinal Study—kindergarten to 8th grade.

Note: Row 8 shows that a male who scores in decile 8 at age 10.5 has a 6.1 percent probability of scoring in decile 5 at age 13.5, a 22.7 percent probability of scoring in decile 8 at age 13.5, and a 10.9 percent probability of scoring in decile 10 at age 13.5. The interpretation of the other cells is analogous, and higher probabilities are associated with darker shading.

eighth grade, and the National Education Longitudinal Study (NELS) which covers people from eighth grade through about age 26 to 27.[10] In each transition matrix, a child is associated with his or her national decile of mathematics skills at a lower age/grade (rows) and again associated with his or her national decile of mathematics skills at a higher age/grade (columns).[11] Thus, if every child stayed in the same decile, 100 percent of the data would fall along the diagonal. Data falling into cells off the diagonal are an indication of a transition—from, say, middling skills in the sixth decile to top skills in the ninth decile.

Figure 1 shows the transition matrix from fifth grade (average age about 10.5) to eighth grade (average age about 13.5) for males. (Males are shown because, although the matrix appears very similar for females, it would be better to start females with fourth grade because of the earlier onset of

10. The two studies are both nationally representative and have many sampling and other methodological choices in common. However, the two studies do not study the same individuals, so there is overlap in the eighth grade but there is not true longitudinal continuity between the two studies. See National Center for Education Statistics, US Department of Education, https://nces.ed.gov/ecls/ and https://nces.ed.gov/surveys/nels88/.

11. I use mathematics rather than verbal tests here, but the verbal results are very similar.

Figure 2. Male Math Score Transitions, Ages 15.5 to 17.5

Deciles of math in grade 10 (age 15.5 on average)	1	2	3	4	5	6	7	8	9	10
1	62.2	28.3	5.7	2.3	0.9	0.2	0.5	0.0	0.0	0.0
2	28.7	43.5	18.3	5.7	2.9	0.8	0.2	0.0	0.0	0.0
3	8.0	26.2	32.7	18.1	9.3	4.0	1.3	0.4	0.0	0.0
4	0.9	8.6	21.9	31.4	20.7	11.8	2.9	2.0	0.0	0.0
5	1.1	3.3	10.4	18.0	27.1	24.1	12.6	3.3	0.0	0.0
6	0.7	1.5	4.0	8.8	20.8	30.5	22.8	9.7	1.3	0.0
7	0.3	1.0	0.8	2.3	9.6	23.9	31.8	24.2	5.6	0.3
8	0.0	0.3	0.4	0.6	2.6	7.9	19.7	37.5	24.8	6.3
9	0.0	0.0	0.0	0.6	1.0	1.3	5.7	21.5	40.7	29.3
10	0.0	0.0	0.0	0.2	0.1	0.5	1.5	5.2	21.6	70.9

Deciles of math in grade 12 (age 17.5 on average)

Source: National Education Longitudinal Study—8th grade onward.

Note: Row 8 shows that a male who scores in decile 8 at age 15.5 has a 2.6 percent probability of scoring in decile 5 at age 17.5, a 37.5 percent probability of scoring in decile 8 at age 17.5, and a 70.9 percent probability of scoring in decile 10 at age 17.5. The interpretation of the other cells is analogous, and higher probabilities are associated with darker shading.

rapid brain development for them. Unfortunately, the ECLS-K does not test children in grade four.) To aid in interpretation, consider row 8 which shows that a male who scores in decile 8 at age 10.5 has a 6.1 percent probability of scoring in decile 5 at age 13.5, a 22.7 percent probability of scoring in decile 8 at age 13.5, and a 10.9 percent probability of scoring in decile 10 at age 13.5. The interpretation of the other cells is analogous, and a cell is shaded darker if the probability shown is higher.[12] In other words, for this example, most of the probability (77.3 percent) is assigned to cells off the diagonal. Students in a middling initial decile have an even higher probability of transitioning out of their initial decile: much of the weight in the center of the transition matrix is off the diagonal. In other words, the matrix suggests that the development of advanced cognitive skills in early adolescence is not wholly predetermined at the beginning of the period.

Figure 2 shows the analogous matrix from tenth grade (average age of 15.5) to twelfth grade (average age of 17.5). Consider row 8. It shows

12. Specifically, the shading is based on intervals of 10 percentage points: 0 to 10 percent, 10.1 to 20 percent, and so on.

that a male who scores in decile 8 at age 15.5 has a 2.6 percent probability of scoring in decile 5 at age 17.5, a 37.5 percent probability of scoring in decile 8 at age 17.5, and a 6.3 percent probability of scoring in decile 10 at age 17.5. Thus, in this example, a student's beginning decile is more determinative: only 62.5 percent of the probability is off the diagonal. Moreover, a student's beginning decile is more determinative regardless of what that decile is. Transitions are less probable everywhere or, to put it another way, skill trajectories are hardening. This hardening or decrease in plasticity is what we would expect from the brain science because the later teenage years are dominated by myelination (accelerating established axons) rather than pruning.

It is worth noting that transition matrixes for younger children, especially very young children, show very little hardening. For instance, an analogous transition matrix for age 1 to 3 would show the vast majority of probability off the diagonal; it is fairly hard even to pick out the diagonal.

II. Mapping Advanced Cognitive Skills in the United States

II.A. Methods and Data

In this section, I map advanced cognitive skills in the United States. While the maps are highly revealing in many ways, they have a few limitations.

First, I have chosen to show a classic US map of counties despite the fact that it overemphasizes sparsely populated counties that have a large land mass. I want viewers to be able to recognize regional patterns and—preferably—even recognize some counties or clusters of counties. For instance, a county that contains a college town might stand out as being prevalent in advanced cognitive skills. Such recognition would not be possible if—for instance—I used a cartogram in which counties' size corresponded to their population.

Second, my measures of advanced cognitive skills depend only on the percentage of people with the skills, not the physical density of people with the skills. That is, I do not show advanced cognitive skills per square mile. This is a deliberate choice because I wish to avoid maps that mechanically predict that rural, low-density areas are skill deserts. Generally speaking, the more sparsely populated a county is, the larger its land mass. It takes self-discipline to ignore vast counties in Nevada, Arizona, Utah, New Mexico, southeastern California, Montana, and the Dakotas.

Third, the only nationally representative measure of cognitive skills among adults in the United States comes from the Program for the International

Table 1. The Distribution of Adults' Scores on Cognitive Skill Tests Administered by the Program for the International Assessment of Adult Competencies

Level	*Below 1*	*1*	*2*	*3*	*4*	*5*
Numeracy (%)	9	20	33	27	9	1
Literacy (%)	4	15	33	35	13	1

Sources: US Department of Education, National Center for Education Statistics, Statistics Canada and OECD, Program for the International Assessment of Adult Competencies (PIAAC), PIAAC 2012–2017 and PIAAC 2017 Literacy, Numeracy, and Problem Solving TRE Assessments.

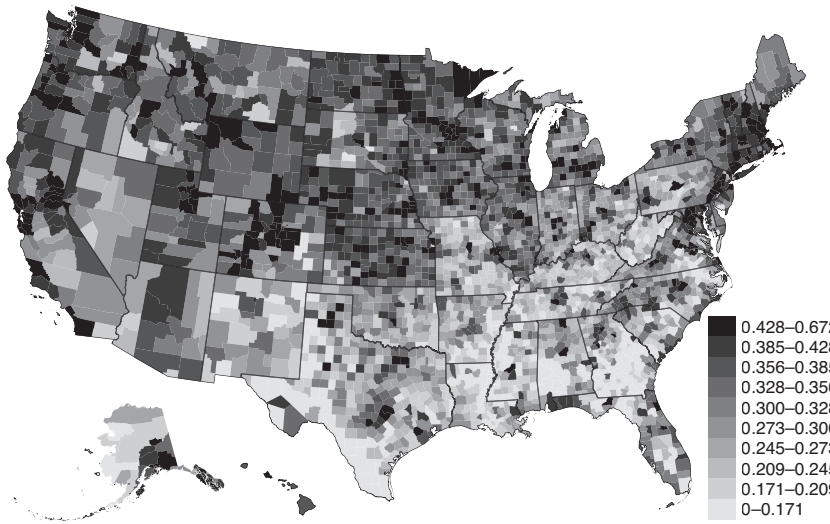
Note: Adults scoring in the top three levels (levels 3–5) can be considered proficient at working with math information and ideas. They have a range of numeracy skills from the ability to recognize math relationships and apply proportions to the ability to understand abstract representations of math concepts and engage in complex reasoning about quantities and data. Totals not equal to 100 percent are due to rounding.

Assessment of Adult Competencies (PIAAC).[13] It measures numeracy (mathematics) and literacy in those age 16 to 65. The PIAAC has six levels of skill for each test, and its tests can identify very advanced skills. For instance, its top level in both numeracy and literacy contains only 1 percent of adults, and its second-to-top numeracy level contains only 9 percent of adults. See table 1. Also, the PIAAC data can in theory be broken down by age, which would be interesting even though it would not be longitudinal data. However, if one wishes to map the data, it is not possible to use them with such fine distinctions (for reasons of confidentiality as well as statistical validity). For mapping, all of the adult ages are consolidated, only county-level statistics are possible, and the PIAAC aggregates the top three levels of skill for each test.[14] Unfortunately, this makes only the numeracy test useful for mapping advanced cognitive skills. Its top three levels contain 37 percent of the adult population—already a coarser definition of "advanced" than I would prefer. However, the literacy test's top three levels contain 49 percent of the adult population, which is simply too coarse. The PIAAC describes adults with the top three levels of numeracy skills as "proficient at working with mathematical information and ideas. They have a range of numeracy skills from the ability to recognize mathematical relationships and apply proportions to the ability to understand

13. "PIAAC Data and Tools," OECD Skills Surveys, https://www.oecd.org/skills/piaac/data.

14. The technical documentation for the county-based statistics from the PIAAC (Krenzke and others 2020) is highly informative and should be consulted by readers interested in how the statistics are computed. Readers interested in the PIAAC more generally should consult Krenzke and others (2019).

Figure 3. Percentage of Adults Whose Numeracy Skills Are at Least Somewhat Advanced (Top Three Levels)



Sources: US Department of Education, National Center for Education Statistics, Program for the International Assessment of Adult Competencies (PIAAC), PIAAC 2012-2017 and PIAAC 2017 Literacy, Numeracy, and Problem Solving TRE Assessments.

Note: Adults scoring in the top three levels can be considered proficient at working with math information and ideas. They have a range of numeracy skills from the ability to recognize math relationships and apply proportions to the ability to understand abstract representations of math concepts and engage in complex reasoning about quantities and data.

abstract representations of mathematical concepts and engage in complex reasoning about quantities and data."[15]

II.B. The Advanced Cognitive Skills Maps

Figure 3 maps advanced cognitive skills by county for the United States, based on the PIAAC numeracy data. The percentage of adults with advanced cognitive skills is divided into ten deciles. In the bottom decile, only 0 to 17 percent of adults have advanced cognitive skills. In the top decile, 42.8 to 67.2 percent of adults have these skills.

15. Institute of Education Sciences, "U.S. PIAAC Skills Map: State and County Indicators of Adult Literacy and Numeracy," https://nces.ed.gov/surveys/piaac/doc/PIAAC-SAE-Brochure.pdf. Sample items from the numeracy test are available at "PIAAC Numeracy—Sample Items," OECD Skills Surveys, https://www.oecd.org/skills/piaac/Numeracy%20Sample%20Items.pdf.

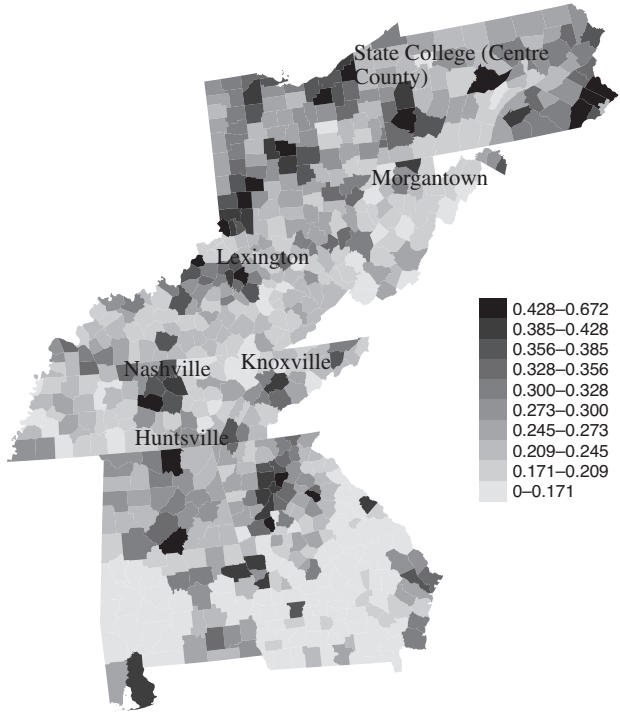
Observe the areas where advanced cognitive skills are most prevalent: most of southern New England but especially the area centered around Boston; scattered counties in New York and New Jersey but especially the area centered around New York City; coastal Washington State and Oregon but especially the area centered around Seattle; coastal California but especially the areas centered around San Francisco, Sacramento, Los Angeles, and San Diego; the Salt Lake City area; and the Washington, DC, area including parts of Maryland and Virginia. None of these findings is surprising. A finding that may surprise some readers is the overall dark shading of what might be called the "Lutheran Belt": southern Minnesota and parts of Wisconsin, Iowa, and the Dakotas.[16] The Lutheran Belt has long had a reputation for strong educational outcomes, and explanations vary. While the area does not contain an advanced cognitive skills concentration like Boston, say, most of its counties fall into the top four deciles of prevalence.

Turning to the other end of the spectrum and the inspiration for this paper's title, where are advanced cognitive skills not prevalent? The most obvious pattern is Appalachia, the area that follows the Appalachian mountain range through several states starting in northeastern Alabama and northwestern Georgia, running though eastern Tennessee and eastern Kentucky, encompassing all of West Virginia, continuing through southeastern Ohio and Pennsylvania, and finally ending in New York.[17] Although Appalachia is the most obvious skill desert, within it there are numerous counties, often in small clusters, in which advanced cognitive skills are prevalent. Figure 4 shows that these are often counties in which a major university is located. Before examining figure 4, however, also note that advanced cognitive skills are not prevalent in the Ozarks (a mountainous plateau mainly in Missouri, Arkansas, and Oklahoma) and in some inland areas of the South—for instance, inland Louisiana, inland Mississippi, and a stretch that begins in north Florida, runs through northeastern Georgia and inland South Carolina, and ends in inland North Carolina.

16. For instance, see the county map of the United States "Lutherans as a Percentage of All Residents, 2000" (ASARB 2002). County percentages based on the total number of adherents reported by the leading Lutheran church bodies, including the Evangelical Lutheran Church in America, the Lutheran Church-Missouri Synod, the Wisconsin Evangelical Lutheran Synod, and the Association of Free Lutheran Congregations, divided by the total population in 2000. The map is available at https://philebersole.wordpress.com/2011/02/22/the-geography-of-american-religion/amp/, accessed October 2021.

17. According to the official listing by the Appalachian Regional Commission, Appalachia also includes some of the western counties of South Carolina, North Carolina, Virginia, and Maryland; "Appalachian Counties Served by ARC," https://www.arc.gov/appalachian-counties-served-by-arc/ (accessed January 2020).

Figure 4. Percentage of Adults in Certain Appalachian States Whose Numeracy Skills Are at Least Somewhat Advanced



Sources: US. Department of Education, National Center for Education Statistics, Program for the International Assessment of Adult Competencies (PIAAC), PIAAC 2012-2017 and PIAAC 2017 Literacy, Numeracy, and Problem Solving TRE Assessments.

Note: Adults scoring in the top three levels can be considered proficient at working with math information and ideas. They have a range of numeracy skills from the ability to recognize math relationships and apply proportions to the ability to understand abstract representations of math concepts and engage in complex reasoning about quantities and data.

I have so far emphasized the most obvious skill deserts and their opposites, areas where advanced cognitive skills are very prevalent. However, many areas of the United States are not easy to characterize, consisting mainly of middling prevalence counties with scattered high prevalence and low prevalence counties.

Figure 4 shows the main counties in Appalachia and picks out a few especially interesting ones that have an unusual degree of advanced cognitive skill relative to the counties surrounding them.[18] All of them are counties in

18. To keep the map readable, I have excluded states where Appalachia runs through only a small part of the state.

which at least one major university exists: the University of Alabama in Huntsville, the University of Tennessee in Knoxville, Tennessee State University and Vanderbilt University in Nashville, the University of Kentucky in Lexington, the University of West Virginia in Morgantown, and Pennsylvania State University in State College. This figure demonstrates that, even if a child grows up in a county where advanced cognitive skills are not prevalent, there is often a county relatively nearby where prevalence is high. Also, the first move to any of these counties might well be to a residential dormitory, presumably a more straightforward move than striking out on one's own in another part of the country.

III. Parsing the Explanations for Advanced Cognitive Skill Deserts

At its outset, this paper reviewed three interesting and non-mutually exclusive explanations for the variation in the prevalence of advanced cognitive skills shown on the maps. The first was very early age factors that affect children through early brain development or other mechanisms that are already fairly determinative at school entry. As noted above, such causes are less likely to directly affect advanced cognitive skills development because the age at which such skills develop is substantially later than ages 0 to 3. Nevertheless, poorly developed noncognitive skills could lead to endogenous responses such as teachers routinely neglecting a child's learning throughout all of the primary grades (kindergarten to grade four) so that when the child reaches early adolescence, he or she is poorly prepared for higher reasoning. For instance, a child might be so confused about rational numbers reasoning, such as multiplication and fractions, that the transition to basic algebra is extremely difficult.

An important reason why it matters how much of a role is played by very early age factors in advanced cognitive skills is that, regardless of how much society might be willing to invest in programs for children between age 0 and 3, custodial care is a meaningful problem. Parents may wish to retain greater time with and control over their children at this stage. In contrast, most adolescents already spend six to eight hours outside the home under the supervision of adults who are not family members: transportation to school, school itself, and organized activities such as sports. Custodial care issues may make very early age factors inherently harder to change through programs conducted by organizations who employ adults other than their parents.

The second explanation was factors that might directly affect the development of advanced cognitive skills because they coincide in timing with the crucial plastic period when these skills begin to form. Obvious examples of such factors would be the middle school teaching or curriculum.

The third explanation was migration toward areas where other people have advanced cognitive skills that are similar to one's own. While it is natural to focus on the migration of individuals who have reached the age of majority (18) and can therefore make location decisions for themselves, selective migration could take place at much earlier ages. For instance, the family of an early adolescent could deliberatively move from an advanced cognitive skill desert to an area where such skills prevail because they believe that the child requires a richer or more demanding school curriculum.

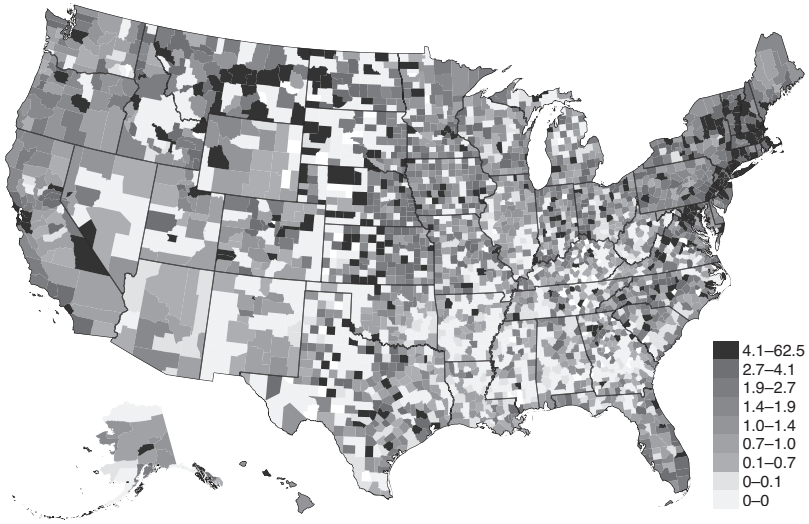
In short, it is helpful to see the age at which the advanced cognitive skills map begins to look like the adult map. Here I work backward, starting from a map based on twelfth graders' SAT and ACT scores, then showing maps based on the test scores of, respectively, eighth graders, fifth graders, and third graders. For these maps, I am constrained to use these test score data; there are virtually no alternatives. Data are needed for all (or at least nearly all) the counties in the United States.[19] Next, I need to be able to put the data for each grade on a common scale across all states. This directs my choice of the third, fifth, and eighth grades, three grades in which all states conduct tests of nearly all their students and for which the states' own test scores can be put on the National Assessment of Educational Progress (NAEP) scale. The NAEP is a test administered to American students in certain grades in a nationally representative sample that is large enough to use for rescaling but far too small for the NAEP's own scores to be mapped.[20] Unfortunately, although the NAEP sample is large enough to be used for rescaling the average score in each county, it is not large enough for computing each county's rescaled percentiles. Thus, I cannot show what I would like most to show: the percentage of children scoring in the advanced cognitive skills range.

The evidence in the maps that follow should be viewed as merely suggestive because there are such significant differences among the data and measures used for the maps at grades three, five, and eight versus grade twelve versus adulthood. It will simply not be possible to state with confidence that the maps' changing patterns only reflect true age-cognition

19. This rules out the survey data gathered by NCES.

20. I use the rescaling provided by the Stanford Education Data Archive (Version 4.1), http://purl.stanford.edu/db586ns4974.

Figure 5. Percentage of Twelfth Graders Whose Math Skills Are at or above the National 90th Percentile



Source: Hoxby and Avery (2013).

Note: The map shows, by decile, the percentage of twelfth graders who scored at or above the national 90th percentile on the math SAT, ACT, PSAT, or pre-ACT. ACT scores were converted to SAT scores using the official 2008 Concordance.

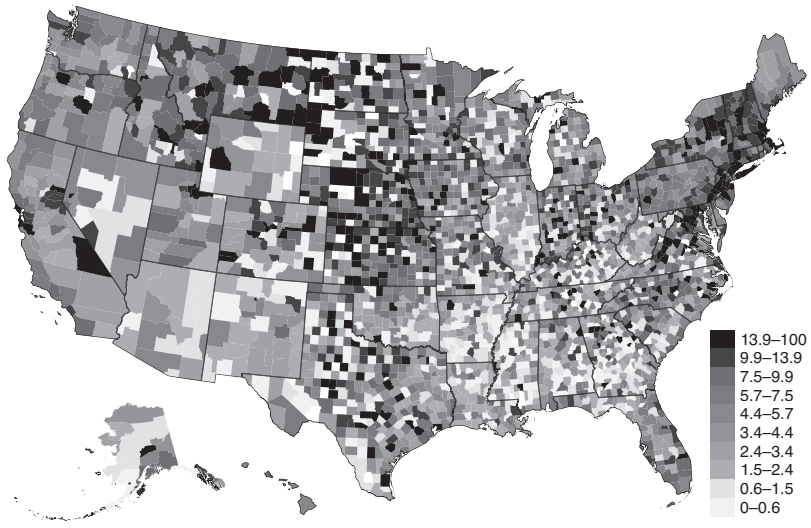
relationships. This weakness motivates the longitudinal empirical exercise that follows the maps.

III.A. Maps of Cognitive Skills for Twelfth Graders, Eighth Graders, Fifth Graders, and Third Graders

Figures 5 and 6 map advanced cognitive skills among twelfth graders. They are based on mathematics SAT and ACT scores and their preliminary versions (PSAT, pre-ACT). Although these tests are not mandatory in all states, students who are likely to score in the advanced range (at or above the 75th percentile) have a high probability of taking at least one of them. ACT scores are converted to the SAT scale based on the official 2008 ACT-SAT Concordance.

In figure 5, which uses the 90th percentile as the threshold for advanced status, the skill deserts are already quite apparent for twelfth graders. Most of the same geographic patterns that appear in the PIAAC-based map also appear in the figure. Figure 6 uses the 75th percentile and, again, the geographic patterns resemble those of the PIAAC-based map. These resemblances

Figure 6. Percentage of Twelfth Graders Whose Math Skills Are at or above the National 75th Percentile



Source: Hoxby and Avery (2013).

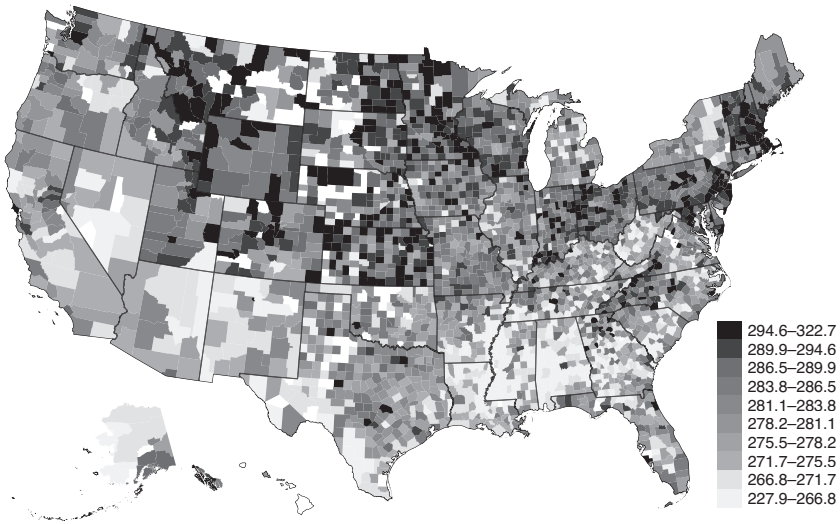
Note: The map shows, by decile, the percentage of twelfth graders who scored at or above the national 75th percentile on the math SAT, ACT, PSAT, or pre-ACT. ACT scores were converted to SAT scores using the official 2008 Concordance.

can be made more concrete. The population-weighted correlation between a county's adult advanced cognitive skills prevalence and its twelfth grade advanced cognitive skills prevalence is 0.56 for the 90th percentile threshold and 0.53 for the 75th percentile threshold.

The map for eighth grade mathematics test scores is shown in figure 7.[21] In the eighth grade, some of the skill deserts that we saw for twelfth graders and adults are already quite apparent: Appalachia and the inland South. This is somewhat striking given that the map is based on average scores, not scores above some threshold like the 75th percentile. (That is, the pattern is expected to shift for this reason alone.) There is also a skill desert on the

21. The test scores are for 2013, the year in which the most complete data are available. For grades five and eight, Virginia does not have mathematics scores that can be rescaled into NAEP scores, owing to its use of end-of-course rather than end-of-grade tests. Therefore, its mathematics scores are predicted for those grades using its English language arts scores, which can be rescaled. If an alternative measure of mathematics (such as the grade-cohort equivalent score, which is available) is used for Virginia, the Virginia map is almost indistinguishable from the one based on predicted NAEP mathematics scores. This suggests that the prediction method gives us a reliable impression for Virginia.

Figure 7. Average Mathematics Skills of Eighth Graders on State Tests (NAEP Scale)



Source: Stanford Education Data Archive.

Note: The map shows, by decile, the average (combined) math and verbal scores of eighth graders of their own state's test. State test scores are rescaled onto the National Assessment of Education Progress (NAEP) scale using the Stanford Education Data Archive, http://purl.stanford.edu/db586ns4974.

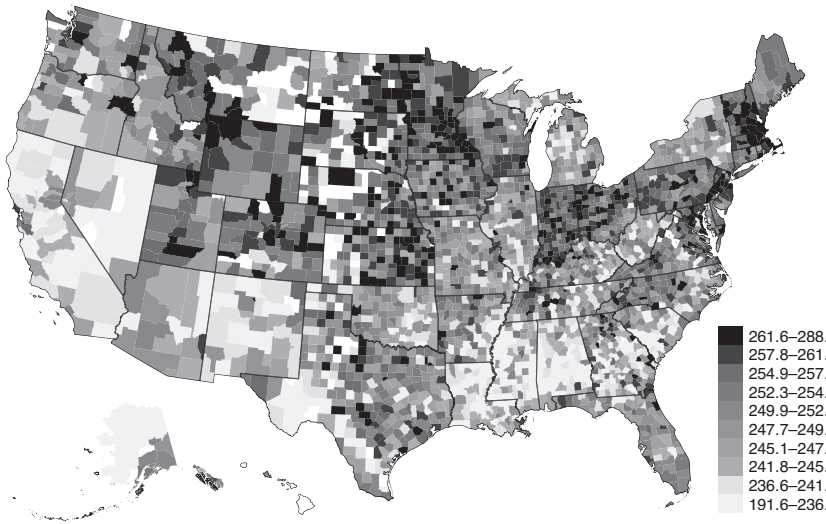
Mexican border that was not nearly so visible on the maps for twelfth graders and adults. One should be cautious with this finding, however, since these counties tend to be sparsely populated though large in land mass. One hypothesis might be that these counties' students are disproportionately likely to be recent immigrants whose lack of English language skills is evident on eighth grade tests but partially made up by the twelfth grade and even more by adulthood.[22] The population-weighted correlation between a county's adult advanced cognitive skills prevalence and its eighth grade cognitive skills prevalence is 0.67.[23]

The analogous maps for fifth and third grade test scores are shown in figures 8 and 9. These grades are important because they represent,

22. English language skills can affect mathematics scores because problems must be read. For instance, when Florida improved its reading scores through a reading-intensive third grade curriculum, its mathematics scores rose as well despite the fact that mathematics lessons were arguably getting less time. See Schwerdt, West, and Winters (2017).

23. The eighth grade measure is based on average scores so that this correlation is not directly comparable to the correlations based on twelfth grade scores where percentile thresholds were used.

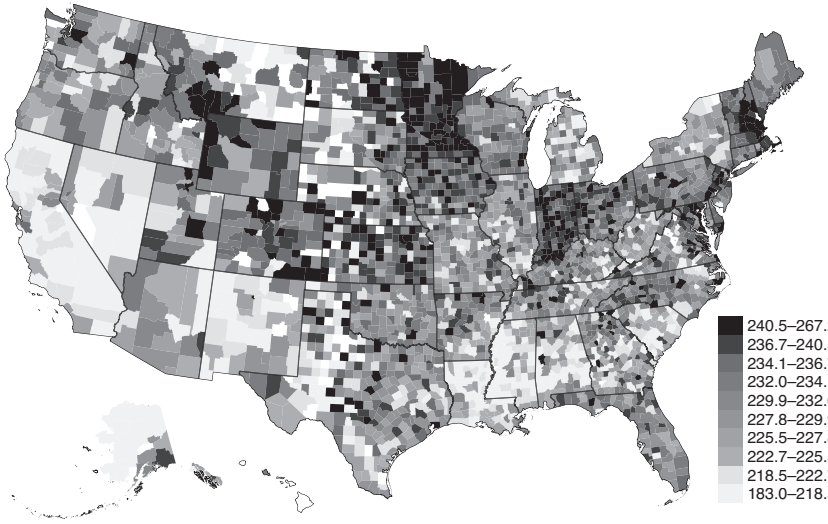
Figure 8. Average Mathematics Skills of Fifth Graders on State Tests (NAEP Scale)



Source: Stanford Education Data Archive.

Note: The map shows, by decile, the average (combined) math and verbal scores of fifth graders of their own state's test. State test scores are rescaled onto the National Assessment of Education Progress (NAEP) scale using the Stanford Education Data Archive, http://purl.stanford.edu/db586ns4974.

Figure 9. Average Mathematics Skills of Third Graders on State Tests (NAEP Scale)



Source: Stanford Education Data Archive.

Note: The map shows, by decile, the average (combined) math and verbal scores of third graders of their own state's test. State test scores are rescaled onto the National Assessment of Education Progress (NAEP) scale using the Stanford Education Data Archive, http://purl.stanford.edu/db586ns4974.

respectively, the very beginning of adolescence and the results of early childhood factors. These maps only somewhat resemble the adult maps. For instance, Appalachia is not nearly so obvious. The pattern across the South is quite different with some coastal areas appearing to be skill deserts while they were not on the twelfth grade or adult maps. Also northern Florida (a skill desert in the twelfth grade and adult maps) shows prevalent cognitive skills among third and fifth graders. Finally, the relative lack of cognitive skills has spread up from the Mexican border counties (eighth grade) deeper into Nevada and California. The population-weighted correlation between a county's adult advanced cognitive skills prevalence and its fifth grade cognitive skills prevalence is 0.54. The analogous number for third grade is 0.50. Notice that these numbers are well below the correlation for eighth grade of 0.67, suggesting that early adolescence is a period when advanced cognitive skill similarities between adults and children is rising rapidly.

What are we to make of these findings? They suggest that a child's cognitive skills in the third grade are not destiny. Even a child's skills in the fifth grade are not destiny. It is only as we look at the map for grade eight that we begin to see patterns that mostly resemble those for adults. By twelfth grade, the resemblance to the adult map is quite strong, though of course imperfect.

III.B. Longitudinal Estimates

One can draw only suggestive evidence by comparing the above maps. Unavoidable differences in data and measures hinder such comparisons. To parse the explanations (very early factors, adolescent factors, migration) with more confidence, longitudinal data on individuals are needed.

The empirical strategy for the exercise is simple.

(1) Regress a child's probability of scoring in the top 30 percent (or 20 percent) on his or her county's share of adults with advanced cognitive skills.

(2) Run this regression for children at each age (N) at which a test score is available.

(3) First run the set of age-specific regressions using adults' advanced cognitive skills from the child's initial county (jt_0) in the longitudinal data.

(4) Plot the regression coefficients to demonstrate how the correlation between children's and adults' cognition changes with the child's age. This evidence should demonstrate the degree to which starting in a county where advanced cognitive skills are more or less prevalent determines the development of a child's cognition.

(5) Repeat the exercise (steps 1 through 4) except this time use adults' advanced cognitive skills from whatever is the child's current county (*jt*) in the longitudinal data. A comparison between the resulting plot and the plot from stage 4 should demonstrate the degree to which migration matters and the age at which important migration tends to occur.

In practice, I combine the no-migration plot from step 4 with the migration-allowed plot from step 5.[24]

The regression estimated for each available age in steps 1 through 2 is:[25]

$$\text{Prob(Child Math Score in Top 30 percent at Age } N)_{ijt} = \alpha_N + \beta_N \text{ Share of Initial County Adults with Advanced Cognitive Skills}_{jt_0} + \varepsilon_{ijt}$$

The analogous regression for each available age in step 5 is:

$$\text{Prob(Child Math Score in Top 30 percent at Age } N)_{ijt} = \alpha_N + \beta_N \text{ Share of Current County Adults with Advanced Cognitive Skills}_{jt} + \varepsilon_{ijt}$$

Although the exercise is simple, the longitudinal data needs are unfortunately challenging. What would be ideal is a longitudinal data set, like the 1970 British Cohort Study, that traces participants from birth through midlife, tracking their cognition and location throughout. No such study exists for the United States, so I use the ECLS-K and NELS. As noted previously, these take us from about age 5.5 to about age 26.5. They both study eighth graders, whose performance looks consistent across the two studies, but they are not truly longitudinal across the eighth grade divide.[26]

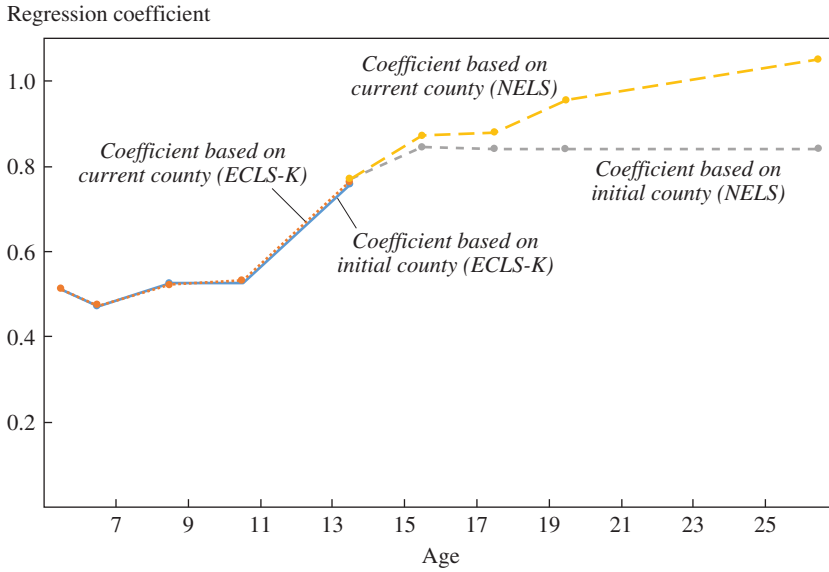
Figure 10 shows the result of the exercise using, as the dependent variable at each age, the probability of a child's math score being in the top 30 percent nationally. The plot for the initial county is quite flat from age 5.5 to 10.5 and indicates a correlation of about 0.5 between a child's cognition and the advanced cognition skills of local adults. The correlation then rises to just above 0.8 at ages 15.5 and 17.5. After that, it appears flat but this is mechanical because no new cognitive measures are available after age 17.5

24. To keep the plots uncluttered, I show only point estimates for the regression coefficients that are plotted. I do not attempt to show the standard errors of the coefficients. In fact, all of the estimated coefficients have *p*-values well below 0.001.

25. I show linear probability coefficients for simplicity, but probit results produce similar findings.

26. I eliminated the National Longitudinal Surveys (https://www.bls.gov/nls/) because their cognitive assessments were too irregular or sparse. I eliminated other NCES longitudinal studies for reasons of timing, follow-up, or the sample frame.

Figure 10. Results of Regressing the Probability of a Child's Math Score Being in the Top 30 Percent on Adults in the County with Math Scores in the Top Third



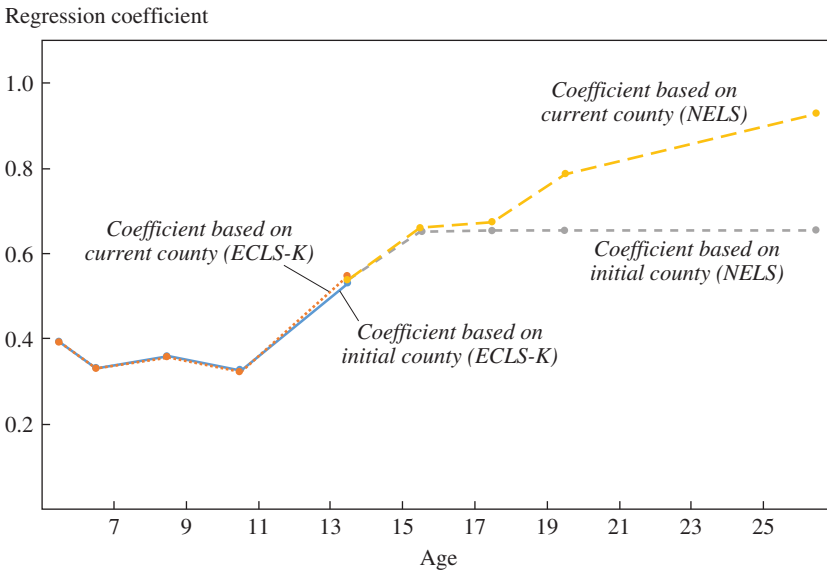
Sources: ECLS-K; NELS; PIAAC.

so that the score from that age is used at ages 19.5 and 26.5 (ages at which we know the person's location but do not have a new cognitive score).

The plot for the current county is very similar to the plot for the initial county between ages 5.5 and 13.5. Between ages 13.5 and 17.5, the plots remain similar although the current line starts to rise slightly relative to the actual line. These results, so far, suggest that consequential migration is insufficiently common before the twelfth grade to affect, more than slightly, the correlations between a child's cognition and the advanced cognitive skills of local adults. That is, migration does occur but it does not greatly alter the prevalence of adults with advanced cognitive skills in the county where a child lives: the counties of origin and destination are sufficiently similar that the coefficients are not affected much.

From age 17.5 onward, migration becomes a substantial factor. The correlation rises from just over 0.8 to more than 0.9 from age 17.5 to age 19.5. In work that is not shown, this appears mainly to be associated with a student's moving to attend college. That is, people who are actually enrolled are at age 19.5 and account for most of the moves from counties with lower

Figure 11. Results of Regressing the Probability of a Child's Math Score Being in the Top 20 Percent on Adults in the County with Math Scores in the Top Third



Sources: ECLS-K; NELS; PIAAC.

advanced cognitive skill prevalence to higher prevalence.[27] Further migration from age 19.5 to age 26.5 raises the correlation still higher—slightly above 1.0 (possible given the method). This is probably mainly associated with job-related moves but the NELS does not contain questions that would allow one to establish this reasonably credibly.

Figure 11 is analogous to figure 10 except that it uses, as the dependent variable at each age, the probability of a child's math score being in the top 20 percent nationally. The coefficient estimate at each age is lower, probably because the top 20 percent threshold used for the ECLS-K and NELS is further from the top 37 percent threshold used for adults' PIAAC scores. Otherwise, the main takeaways are largely the same: the correlation between a child's and local adults' cognition starts at a modest level (about 0.35), rises between the ages of 10.5 and 15.5, and is flat thereafter

27. College-related moves are not always stereotypical moves to a university town. There are numerous moves to, say, an adjoining county where a community college is available.

unless one takes account of migration. Migration between ages 17.5 and 26.5 accounts for a substantial further rise in the correlation.

Though causality is not established, the evidence from the longitudinal exercise suggests that local adults' advanced cognitive skills probably have some effect on a child's advanced cognitive skills starting at an early age, and this influence appears to rise starting in early adolescence. By saying that causality is not established, I mean that local adults' cognitive skills need not be the channel through which the child is affected. The channel could be any variable(s) systematically correlated with the prevalence of advanced cognitive skills among adults: schooling variables, employment variables, income variables, health variables, and so on. I return to this point in the discussion.

Migration after high school accounts for a continuing increase in the correlation but this is almost certainly due mainly to self-selective moves. I have presented no evidence, as from a random experiment, showing that people randomly induced to move to a county where advanced cognitive skills are more prevalent see a consequent rise in their own skills.[28] Studies that rely on nonexperimental family moves, especially those that substantially change a child's environment, do not produce credibly causal estimates. Major family moves are not made in a random or trifling way: they are simply too expensive and potentially consequential.

IV. Correlates of Advanced Cognitive Skills

As mentioned at the outset, this paper does not attempt causal tests of explanations why advanced cognitive skills are more prevalent in some areas than in others. However, by examining some correlates of prevalence, it may be possible to focus on some explanations. Keep in mind, though, that a variable that is highly correlated with prevalence may be so because of reverse causality. For instance, students with low cognitive skills may find it hard to succeed in college.

Table 2 shows county-level correlations between the advanced cognitive skill measures mapped above and a variety of socioeconomic factors.[29]

The first eight rows show correlations with closely related measures of cognitive skill. One takeaway is that the prevalence of advanced cognitive skills is highly negatively correlated with the prevalence of adults who

28. For a credible natural experiment along these lines, see Kawano and others (2018).
29. The correlations are weighted by the population age 8 to 65.

Table 2. Correlations between Advanced Cognitive Skill Measures and Other Factors

	Adults with advanced numeracy skills (%)	*Twelfth graders with math skills ≥ 90th percentile*	*Twelfth graders with math skills ≥ 75th percentile*	*Average math skills of eighth graders*	*Average math skills of fifth graders*	*Average math skills of third graders*
Adults with low numeracy skills (%)	−0.86	−0.44	−0.49	−0.73	−0.67	−0.66
Adults with low literacy skills (%)	−0.76	−0.36	−0.42	−0.68	−0.63	−0.63
Adults with advanced literacy skills (%)	0.99	0.56	0.55	0.67	0.60	0.56
Twelfth graders with verbal skills ≥ 75th percentile	0.51	0.91	0.97	0.49	0.50	0.47
Twelfth graders with verbal skills ≥ 90th percentile	0.55	0.95	0.96	0.50	0.50	0.44
Average English language arts skills of eighth graders	0.67	0.58	0.60	0.90	0.82	0.76
Average English language arts skills of fifth graders	0.62	0.55	0.58	0.82	0.88	0.84
Average English language arts skills of third graders	0.54	0.51	0.55	0.75	0.82	0.87
Adults with less than high school (%)	−0.78	−0.42	−0.38	−0.60	−0.58	−0.57
Adults with high school or some college less than a BA (%)	−0.44	−0.42	−0.57	−0.27	−0.17	−0.10
Adults with BA or MA (%)	0.83	0.60	0.69	0.53	0.50	0.43
Adults with professional or doctoral degree (%)	0.63	0.48	0.61	0.29	0.27	0.22
Public school spending per student	0.22	0.39	0.34	0.19	0.13	0.08
Public school staff compensation per student	0.24	0.43	0.40	0.19	0.15	0.11
Public school instructor wages	0.26	0.43	0.40	0.21	0.19	0.15
Students in private schools (%)	0.35	0.38	0.33	0.14	0.21	0.19
Pre-K student–teacher ratio	−0.12	−0.06	−0.08	−0.13	−0.17	−0.19
Kindergarten student–teacher ratio	0.00	−0.03	−0.04	−0.03	−0.02	−0.04
Primary school student–teacher ratio	0.19	−0.07	−0.07	0.05	0.05	0.05
Middle school student–teacher ratio	0.07	−0.16	−0.14	−0.07	0.00	0.05
High school student–teacher ratio	−0.02	−0.05	−0.05	−0.07	−0.07	−0.07
Poverty rate	−0.61	−0.52	−0.56	−0.64	−0.58	−0.54
Disability rate	−0.59	−0.52	−0.45	−0.54	−0.41	−0.33
Households on food stamps (%)	−0.68	−0.51	−0.52	−0.71	−0.55	−0.50
Unemployed or out of the labor force (%)	−0.63	−0.40	−0.41	−0.53	−0.55	−0.51
Child households not headed by husband and wife (%)	−0.64	−0.48	−0.51	−0.77	−0.68	−0.63
Households rural (%)	−0.34	−0.29	−0.19	−0.14	−0.08	0.01

Sources: PIAAC, Hoxby and Avery (2013), SEDA, Manson and others (2021), IPUMS National Historical Geographic Information System 2010 Census and 2000–2012 American Community Survey data, NCES Common Core of Data 2010–2011 files.

Note: Table shows population-weighted, county-level correlations between measures of cognitive skill and other factors.

score at the lowest levels (below level 1 and level 1) on the PIAAC exam. Indeed, a map of low cognitive skills looks much like a reverse image of the map of high cognitive skills. Another takeaway is that the math-based measures are highly correlated with parallel measures based on verbal skills. In short, little evidence has been lost by focusing on math-based, advanced skill measures.

The next four rows show correlations with educational attainment. Advanced cognitive skills are negatively correlated with the share of the population with no high school degree or with a high school degree and some college but less than a baccalaureate degree. Advanced cognitive skills are positively correlated with the share of the population with a baccalaureate or master's degree or with a professional or doctoral degree. These correlations are stronger when advanced cognitive skills are measured later in life, confirming some of the evidence presented above that early cognitive skills are not as determinative. Reverse causality is a serious issue for these correlations, but they are a useful sanity check.

The next nine rows show correlations with measures of the K-12 educational experience. Advanced cognitive skills in the twelfth grade and adulthood are positively correlated with measures of school spending, staff compensation, and the share of students who attend private school. The correlations weaken as one moves to cognitive skill measures based on earlier ages. Of course, these correlations may merely be picking up income since it often affects school spending and families' ability to pay tuition. Advanced cognitive skills have little evident correlation with the student-teacher ratio at various grades. Even the signs are not consistently negative as might be expected. This may seem surprising, but it is a finding that is standard in the correlational literature on class size and a finding also of at least one credibly causal study.[30]

The final six rows show correlations with the poverty rate, disability rate, share of households who receive food stamps (SNAP), share of adults who are unemployed or out of the labor force, share of households with children that are not headed by a husband and wife, and share of households that are rural. Advanced cognitive skills are negatively correlated with each of these measures. Although not shown, advanced cognitive skills are positively correlated with median and mean household income.

It is worth noting that advanced cognitive skills have little or no correlation with a variety of measures of housing (such as the percentage who rent

30. See Hanushek (1997) for a review of mainly correlational studies. See Hoxby (2000) for a credibly causal (regression discontinuity) study that finds negligible effects of class size.

or the vacancy rate) or with a variety of measures of age distribution (such as the percentage under age 18 or over age 65).

V. Discussion

V.A. Implications for Policies Likely to Affect Children Directly

I have presented evidence that there is substantial variation in the prevalence of advanced cognitive skills and that skill deserts exist where they are far from prevalent. The evidence suggests that skill deserts have factors that affect adolescents disproportionately and are associated with a local absence of adults with advanced cognitive skills. These two conditions suggest numerous possibilities. For instance, in a skill desert, there could be shortages of teachers who have the skills to teach rich and demanding curricula. There could be obstacles to putting such curricula into schools, even if skilled teachers are available. Obstacles might include a lack of funding for schools or low population density that makes it hard for schools to be of sufficient size for specialized classes. There could be a lack of sophistication among adults about college-going so that adolescents do not realize how important it is to acquire advanced cognitive skills to succeed in higher education—many two-year degree programs as well as baccalaureate programs.

The possibilities mentioned here are all concerned with education, and this makes sense if we focus on frontal lobe development. However, the frontal lobe also has a disproportionate effect on planning and self-regulation. In addition, puberty begins in early adolescence. So there are candidate explanations that fit the criteria but are not about learning per se. For instance, if cultural norms in an area expect teens to become parents, then students might tune out school once they hit puberty and adolescence. Thus, purely through fertility expectations, a child growing up in a county with low prevalence of advanced cognitive skills among adults could be less likely to develop such skills. Furthermore, this relationship would intensify as puberty hit. This story would fit the conditions and yet have no direct link to education. Solving a shortage of skilled teachers might be entirely ineffectual. In this scenario, education is merely collateral damage.

Other stories characterized by omitted variables are possible. Although there are numerous studies of cognitive skill development, only a small fraction meets standards of causality. Those that do tend to be randomized controlled trials or policy experiments that affect a limited number of students in a limited range of grades. For a review of adolescent-oriented studies that classifies them according to standards of causality, see Herrera,

Truckenmiller, and Foorman (2016). Large-scale natural experiments are likely needed if we are to learn more about how the same treatment, such as teacher quality, might have different effects at different ages. These are what I employ in the Tanner Lectures mentioned above.

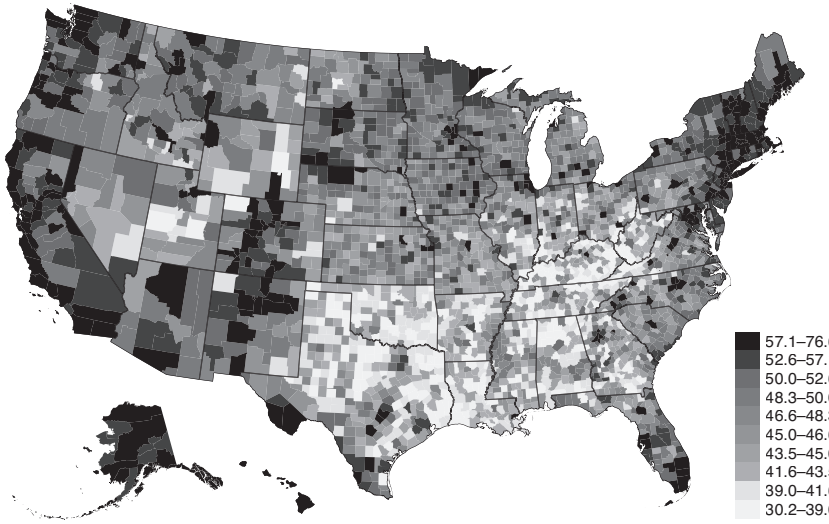
What we can almost certainly say, even without knowing the exact causes, is that we need greater analysis of the early adolescent period of schooling and brain development. In particular, since the correlation between a twelfth grader's cognitive skills and local prevalence of advanced cognitive skills among adults is still well below one, a child who begins life in a skill desert nevertheless has a reasonable probability of attaining advanced cognitive skills. More work is needed to understand why some adolescents are more resilient to such an environment. Adolescence does indeed appear to be an "age of opportunity," as expressed by so many researchers who study brain development.

V.B. Speculative Implications for Policies Unlikely to Affect Children Directly

In the process of writing this paper, I began to speculate that some children born in advanced cognitive skill deserts might reasonably develop economic fatalism, resentment toward perceived intellectual elites, or a lack of social trust in people whose skills differ from the adults most familiar to them. If so, there may be additional urgency in understanding geographic variation in the prevalence of advanced cognitive skills. Put another way, there may be additional urgency in making adolescence an age of opportunity rather than an age of risk.

Charles, Hurst, and Schwartz (2019) show that when a local area experiences a decline in manufacturing, there are large and—more importantly for economic fatalism—persistent negative effects on local employment rates, hours worked, and wages. Since jobs in the declining fields (manufacturing, mining) are less likely to require advanced cognitive skills than jobs in rapidly growing fields (technology, health care), students who reach the end of high school without acquiring advanced cognitive skills may realistically foresee a long future ahead with little hope of achieving economic security. Indications of such economic fatalism come from Charles, Hurst, and Schwartz (2019), who show associations between declining local manufacturing and rising local opioid use and deaths. Case and Deaton's (2015, 2017, 2020) influential work has also associated "deaths of despair" with economic fatalism. They have mapped such deaths, and readers can compare their maps to the maps of advanced cognitive skills in this paper. Some notable similarities will be observed.

Figure 12. Estimated Share of Adults Who Believe Most Scientists Think Global Warming Is Happening



Source: Howe and others (2015).

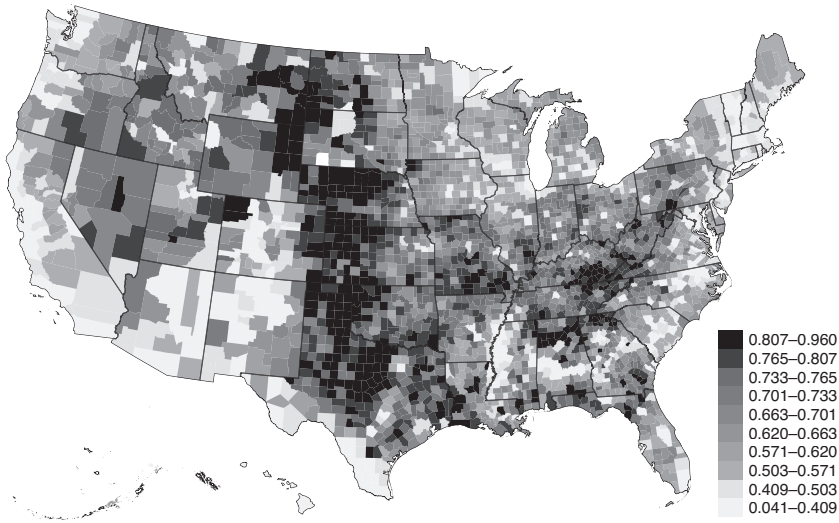
If people perceive technical change to be skill-biased (and persistently skill-biased for the foreseeable future), then those who do not learn advanced cognitive skills in adolescence may expect that demand for their skills will steadily fall while demand for the skills of perceived intellectual elites will steadily rise. This could easily lead to resentment of or distrust in the perceived elites.

Finally, if global trade and immigration policies are perceived to be exacerbating the economic forces, such as skill-biased technical change, that favor those with advanced cognitive skills, then students who reach the end of high school without acquiring advanced cognitive skills may align themselves against such policies. This might lead them to vote for politicians who oppose trade and/or immigration.

Figure 12 shows the US county map of the estimated share of adults who believe "most scientists think global warming is happening."[31] Note that this is not a question about whether the adult himself or herself believes in global warming, but a question about what most scientists think. This

31. Quoted from the question wording in the Climate Change in the American Mind Survey, a tab on the Yale Climate Opinion Maps by Howe and others (2015), accessed January 2020. This map is based on data from Howe and others (2015).

Figure 13. Share of Votes for Donald J. Trump in 2016



Source: MIT Election Data and Science Lab, 2018, "County Presidential Election Returns 2000– 2016," Harvard Dataverse, V6, https://doi.org/10.7910/DVN/VOQCHQ.

question was selected because the answer (yes) can be easily established and is well known.[32]

Figure 13 shows the US county map of the share of votes for Donald J. Trump in 2016.[33] His platform was notably more protectionist and anti-immigration than those of his opponents, including most of his primary opponents.

ACKNOWLEDGMENTS The author gratefully acknowledges comments from Janice Eberly, Erik Hurst, Brian Jacob, and James Stock.

32. The consensus is that 90 percent or more of scientists think global warming is happening. While this number could be wrong, the true number is very unlikely to be lower than or equal to 49 percent. See Cook and others (2016).

33. This map is based on data from the MIT Election Data and Science Lab, 2018, "County Presidential Election Returns 2000–2016," Harvard Dataverse, V6. https://doi.org/10.7910/DVN/VOQCHQ, accessed January 2020.

References

Association of Statisticians of American Religious Bodies (ASARB). 2002. *Religious Congregations and Membership in the United States, 2000*. Nashville: Glenmary Research Center.

Atkins, S. M., M. F. Bunting, D. J. Bolger, and M. R. Dougherty. 2012. "Training the Adolescent Brain: Neural Plasticity and the Acquisition of Cognitive Abilities." In *The Adolescent Brain: Learning, Reasoning, and Decision Making*, edited by Valerie F. Reyna, Sandra B. Chapman, Michael R. Dougherty, and Jere Confrey. Washington: American Psychological Association.

Case, Anne, and Angus Deaton. 2015. "Rising Morbidity and Mortality in Midlife among White Non-Hispanic Americans in the 21st Century." *Proceedings of the National Academy of Sciences* 112, no. 49: 15078–83.

Case, Anne, and Angus Deaton. 2017. "Mortality and Morbidity in the 21st Century." *Brookings Papers on Economic Activity*, Spring, 397–443.

Case, Anne, and Angus Deaton. 2020. *Deaths of Despair and the Future of Capitalism*. Princeton, N.J.: Princeton University Press.

Chapman, Sandra B., Jacquelyn F. Gamino, and Raksha Anand Mudar. 2012. "Higher Order Strategic Gist Reasoning in Adolescence." In *The Adolescent Brain: Learning, Reasoning, and Decision Making*, edited by Valerie F. Reyna, Sandra B. Chapman, Michael R. Dougherty, and Jere Confrey. Washington: American Psychological Association.

Charles, Kerwin Kofi, Erik Hurst, and Mariel Schwartz. 2019. "The Transformation of Manufacturing and the Decline in US Employment." *NBER Macroeconomics Annual* 33:307–72.

Cook, John, Naomi Oreskes, Peter T. Doran, William R. L. Anderegg, Bart Verheggen, Ed W. Maibach, and others. 2016. "Consensus on Consensus: A Synthesis of Consensus Estimates on Human-Caused Global Warming." *Environmental Research Letters* 11, no. 4: 048002.

Couture, Victor, Cecile Gaubert, Jessie Handbury, and Eric Hurst. 2020. "Income Growth and the Distributional Effects of Urban Spatial Sorting." Working Paper 26142. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26142.

Dahl, Ronald E. 2004. "Adolescent Brain Development: A Period of Vulnerabilities and Opportunities." *Annals of the New York Academy of Sciences* 1021:1–22.

Forman, Chris. 2013. "How Has Information Technology Use Shaped the Geography of Economic Activity?" In *Handbook of Industry Studies and Economic Geography*, edited by Frank Giarratani, Geoffrey J. D. Hewings, and Philip McCann. Northampton, Mass.: Edward Elgar Publishing.

Hanushek, Eric A. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis* 19, no. 2:141–64.

Herrera, S., A. J. Truckenmiller, and B. R. Foorman. 2016. *Summary of 20 Years of Research on the Effectiveness of Adolescent Literacy Programs and Practices*. Washington: US Department of Education, Institute of Education Sciences,

National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://ies.ed.gov/ncee/edlabs.

Howe, Peter D., Matto Mildenberger, Jennifer R. Marlon, and Anthony Leiserowitz. 2015. "Geographic Variation in Opinions on Climate Change at State and Local Scales in the USA." *Nature Climate Change* 5:596–603. https://doi.org/10.1038/nclimate2583.

Hoxby, Caroline M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics* 115, no. 4: 1239–85.

Hoxby, Caroline M., and Christopher Avery. 2013. "The Missing 'One-Offs': The Hidden Supply of High-Achieving, Low-Income Students." *Brookings Papers on Economic Activity*, Spring, 1–50.

Kawano, Laura, Bruce Sacerdote, Mike Stevens, and William Skimmyhorn. 2018. "The Determinants of Children's Outcomes: Evidence from Random Geographic Movement of Military Families." Working Paper. https://conference.nber.org/conf_papers/f95269.pdf.

Kolko, Jed. 2007. "Agglomeration and Co-Agglomeration of Services Industries." Social Science Research Network, May 16. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=985711.

Krenzke, T., L. Mohadjer, J. Li, A. Erciulescu, R. Fay, W. Ren, W. Van de Kerckhove, L. Li, and J. N. K. Rao. 2020. *Program for the International Assessment of Adult Competencies (PIAAC): State and County Estimation Methodology Report (NCES 2020-225)*. Washington: US Government Printing Office. Retrieved December 2020 from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020225.

Krenzke, T., W. VanDeKerckhove, N. Thornton, L. Diaz-Hoffmann, J. Hogan, L. Mohadjer, and others. 2019. *U.S. Program for the International Assessment of Adult Competencies (PIAAC) 2012/2014/2017: Main Study, National Supplement, and PIAAC 2017 Technical Report (NCES 2020-224)*. Washington: National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020224.

Malamud, Ofer, and Abigail Wozniak. 2012. "The Impact of College on Migration: Evidence from the Vietnam Generation." *Journal of Human Resources* 47, no. 4: 913–50.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2021. *IPUMS National Historical Geographic Information System: Version 16.0* [dataset]. Minneapolis: University of Minnesota. http://doi.org/10.18128/D050.V16.0.

Mori, Tomoya. 2017. "Agglomeration." Working Paper 960. Kyoto: Kyoto University, Institute of Economic Research.

Piaget, Jean. 1972. "Intellectual Evolution from Adolescence to Adulthood." *Human Development* 15: 1–12.

Reyna, Valerie F., Sandra B. Chapman, Michael R. Dougherty, and Jere Confrey. 2012. "Introduction." In *The Adolescent Brain: Learning, Reasoning, and Decision Making*, edited by Valerie F. Reyna, Sandra B. Chapman, Michael R. Dougherty, and Jere Confrey. Washington: American Psychological Association.

Schwerdt, Guido, Martin R. West, and Marcus A. Winters. 2017. "The Effects of Test-Based Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida." *Journal of Public Economics* 152:154–69.

Spear, Linda Patia. 2013. "Adolescent Neurodevelopment." *Journal of Adolescent Health* 52, no. 2 Suppl.: S7–13.

Steinberg, Laurence D. 2015. *Age of Opportunity: Lessons from the New Science of Adolescence*. Boston: Houghton Mifflin.

Comments and Discussion

COMMENT BY

ERIK HURST This paper documents a very interesting set of facts. In particular, the paper shows spatial variation in the extent to which adult residents have a deficit in what Hoxby defines as "advanced cognitive skills." The paper then speculates that advanced cognitive skill deserts may arise because of differential local investments in these skills during adolescence. Overall, I expect that the findings in this paper will stimulate a large amount of future research.

The paper is mostly descriptive. As a result, I do not have many substantive comments on the paper's message. However, three things entered my mind as I was reading it. First, I wondered about other ways to measure advanced cognitive skills within a local area. Second, I wondered whether information on within-county variation in cognitive skills would complement the paper's cross-county analysis. Finally, I wondered whether other correlates may be useful to readers with respect to understanding the causes of advanced cognitive skill deserts. I expand on each of these comments below.

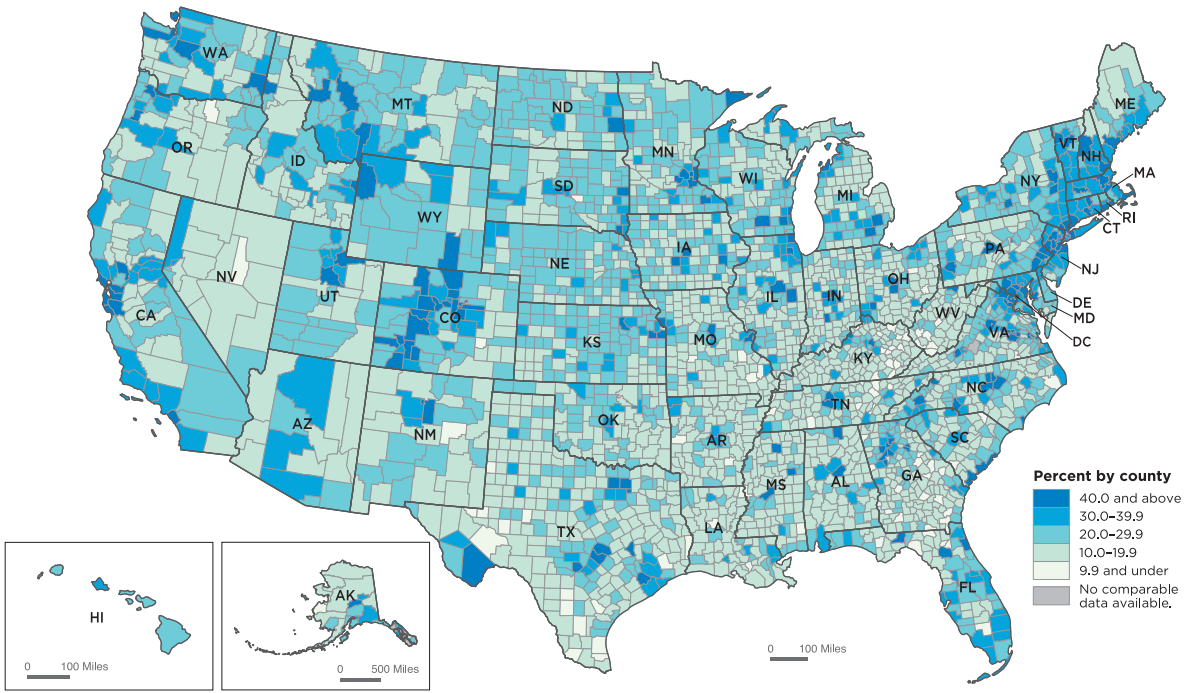
HOW TO MEASURE COGNITIVE SKILL DESERTS In the paper, advanced cognitive skills are defined broadly as those skills needed to perform higher order reasoning. In particular, the paper refers to advanced cognitive skills as those that "require a capacity to solve problems through logic, think in the abstract, engage in critical thinking, and derive general principles from a set of facts." To create maps of cognitive skill deserts, the paper uses data from the Program for the International Assessment of Adult Competencies (PIAAC). The PIAAC is a large survey that assesses respondents' numeracy and literacy skills. Respondents are binned into six levels of skills after taking each test. The paper defines respondents as having advanced cognitive skills if they are classified in the three highest levels of skill on

the numeracy portion of the PIAAC. By this definition, about 37 percent of the US adult population is classified as having advanced cognitive skills. Figure 3 of the paper shows the share of respondents in each county who have advanced cognitive skills by this metric. The cognitive skill deserts documented in the paper using this metric are concentrated almost exclusively in the South census region. College towns throughout the South show higher levels of advanced cognitive skills. But, for the most part, the rest of the South—including much of Appalachia—has more counties with low levels of advanced cognitive skills, particularly compared to other regions.

The US Census puts out maps showing variation across counties in the share of residents over the age of 25 with a bachelor's degree or higher (McElrath and Martin 2021). In my discussion, I showed one of these maps (figure 1). The Census Bureau creates this map using individual-level data pooling together the 2015–2019 waves of the American Community Survey. The patterns of the spatial variation in the share of residents with at least a bachelor's degree is nearly identical to the spatial variation in the advanced cognitive skills as measured by the PIAAC documented in the paper. In particular, in the United States there is a "bachelor's degree desert" in the South. Within the South, there are pockets of counties with a higher share of bachelor's degrees. These counties often include college towns and are the same counties in the South that have a larger amount of residents with advanced cognitive skills. The similarity in spatial patterns across the two measures begs the question of whether cognitive skill measures from the PIAAC are just proxying for lower levels of accumulated schooling. It would have been nice to have a plot in the paper correlating county-level share of high cognitive skills using the PIAAC numeracy measures with the county-level share of residents with a bachelor's degree. Are the numeracy measures proxying for something above and beyond low levels of accumulated education? Are the cognitive skill deserts highlighted in the paper simply places where education levels of adult residents are low? What are high scores on numeracy exams from the PIAAC measuring that is distinct from obtaining a bachelor's degree? The paper is silent on these questions. Going forward, it may be useful to flush out whether cognitive skill deserts are something distinct relative to places with lower levels of accumulated schooling.

WITHIN-COUNTY VERSUS CROSS-COUNTY VARIATION IN ADVANCED COGNITIVE SKILLS My second comment on the paper centers on what is the correct level of aggregation to think about cognitive skill deserts. The paper uses variation across counties. In doing so, it suggests that certain counties may invest less in developing advanced cognitive skills relative to other counties.

Figure 1. Percentage of People 25 Years and Older with a Bachelor's Degree or Higher, 2015–2019



Sources: US Census Bureau, table S1501, 2010–2014 American Community Survey, five-year estimates; https://www.census.gov/content/dam/Census/library/publications/2021/acs/acsbr-009.pdf. Used with permission.

Notes: Due to the county boundary changes that occurred within the 2005–2009 and 2010–2019 time periods, the following county equivalents appear in the "No comparable data available" category: Petersburg, Alaska; Prince of Wales-Hyder, Alaska; Hoonah-Angoon, Alaska; Bedford, Virginia; and Bedford City, Virginia.

It would be interesting to think about within-county variation as well. Consider, for example, the city of Chicago. Chicago is comprised of dozens of different neigbhorhoods. I would conjecture that the variation across neighborhoods within Chicago with respect to the paper's measure of advanced cognitive skills is as large as the variation across counties. There is some evidence to back this up. The *Chicago Tribune* (2017) reports average student SAT scores by Chicago high school. The lowest average SAT math score from students in a Chicago high school was 360 while the highest average SAT math score was 686. The schools with lower SAT math scores were in geographically different areas within Chicago than the schools with higher scores. There are parts of Chicago that would look like they were in a cognitive skill desert (measured by SAT math scores) relative to other parts of Chicago.

Going forward, it may be useful to explore the extent to which within-county variation in measures of cognitive skills are useful in helping us learn about the causes of cross-county variation in measures of cognitive skills. There is large spatial variation in measures of schooling or test scores even within a large city like Chicago. Why is it more interesting to focus on cross-county differences in cognitive skills relative to focusing on within-county spatial differences? Future work can shed light on these issues.

WHAT EXPLAINS THE EXISTENCE OF COGNITIVE SKILL DESERTS? My third and final comment centers on potential explanations for the spatial variation in advanced cognitive skills. The paper focuses on a handful of potential explanations for the cognitive skill deserts. The first discusses early childhood factors, and the second focuses on influences during adolescence. The paper shows that advanced cognitive skills measures of adults (the PIAAC data) in a given county correlate strongly with test score measures of adolescents in that location. However, the adult measures of cognitive skills in a given location are only weakly correlated with test score measures of younger children. The paper then concludes that advanced cognitive skills are mostly engrained in adolescence. That conclusion rests on the extent to which test scores of young children actually measure well a child's cognitive skills. If test scores measure cognitive skills with error and that error is larger for younger children than for older childer, we would expect more spatial correlation between the test scores of adults and the test scores of teenagers than we would between the test scores of adults and the test scores of younger children.

Going forward, it would be useful to explore other demographic and socioeconomic correlates of spatial differences in measures of skills. For example, how does spatial variation in PIAAC cognitive skill measures vary

with spatial differences in parental education, adult income, adult industry mix, and other demographic variables (such as race and ethnicity)? These correlations can help shed light on some of the mechanisms underlying the spatial variation in measures of skills.

REFERENCES FOR THE HURST COMMENT

Chicago Tribune. 2017. "SAT Scores in Illinois: See How Your High School Compares." November 9. https://www.chicagotribune.com/suburbs/ct-school-report-card-sat-scores-2017-htmlstory.html.

McElrath, K., and M. Martin. 2021. "Bachelor's Degree Attainment in the United States: 2005 to 2019." American Community Survey Brief ACSCR-009. Washington: US Department of Commerce, US Census Bureau. https://www.census.gov/content/dam/Census/library/publications/2021/acs/acsbr-009.pdf.

COMMENT BY

BRIAN A. JACOB In this paper Hoxby examines the variation in cognitive ability across geographic locations in the United States. She documents three important facts. First, adults with advanced cognitive skills are clustered disproportionately in certain places. Specifically, urban and coastal areas have a particularly high proportion of adults with advanced skills. Examples include northern New England (such as Boston), large metropolitan areas in California, and selected counties in the Midwest that Hoxby refers to as the "Lutheran Belt." Appalachia, the Ozarks, and areas of the inland South (Louisiana, Mississippi, Georgia) are "skill deserts," with very few adults possessing advanced cognitive skills.

Second, cognitive skills among children are distributed more evenly across geography compared with cognitive ability of adults. While children in urban and coastal areas and parts of the Midwest tend to outperform those in Appalachia and the Southeast, the differences are much less stark than in adults.

Third, there is a correlation between the geographic distribution of adult skills and the analogous distribution of child skills. Importantly, the magnitude of the correlation increases as children get older, particularly as they enter adolescence. That is, the correlation between adult skills and the achievement level of high school students is larger than the correlation between adult skills and elementary school achievement.

Hoxby argues that these facts, in combination with other evidence, have important implications. Throughout the paper, she emphasizes two related themes. One involves the salience of advanced cognitive skills. She first argues that skill-biased technological change and related economic forces

have increased the importance of such skills in today's labor market. The second theme is the importance of adolescence as an "age of opportunity." Referencing brain science research indicating that adolescence is the time during which advanced cognitive skills develop, she suggests that it might be particularly beneficial to target educational interventions during adolescence. Based on the geographic skill distribution, policymakers should target skill deserts in particular for such interventions.

There is a lot to like in Hoxby's analysis. First, the attention on adolescence is a useful antidote to the policy community's intense focus on early childhood over the past two decades.[1] This is not to say that educators should avoid intervening early in children's lives, but rather that the intense focus on this time period risks neglecting effective strategies for older children. Second, the focus on geography is consistent with other recent work, such as the analysis of intergenerational mobility by Chetty and Hendren (2018a, 2018b) and Chetty, Hendren, and Katz (2016). In particular, Hoxby highlights the challenges faced by rural communities, which are sometimes neglected as policymakers have focused (understandably) on the struggles of those in urban areas.

In this comment, I seek to make several points. To begin, I raise some methodological issues that complicate Hoxby's analysis. Second, I provide some supplementary evidence to support Hoxby's contention that students in the United States struggle during adolescence. Finally, I discuss what evidence we have on potential interventions for adolescents in skill deserts.

THE CHALLENGE OF ASSESSING AND INTERPRETING COGNITIVE ABILITY I wholeheartedly agree with Hoxby's contention that cognitive skills are a more useful measure of an individual's capacity to function in contemporary labor markets than educational attainment. Unfortunately, assessing individual skills presents a number of challenges.[2]

First, standardized test scores are noisy measures of true ability, which fluctuate for many reasons, ranging from whether an individual was sick or distracted during the test to which particular items were asked on the assessment. Moreover, there are reasons to believe that the degree of measurement error may vary based on factors such as the age or gender of the test taker.

1. As an example, enrollment in state-funded prekindergarten programs has risen dramatically in the past two decades, from roughly 3 percent (14 percent) of three-year-olds (four-year-olds) in 2002 to 6 percent (34 percent) in 2020. See Friedman-Krauss and others (2021, 9).

2. Jacob and Rothstein (2016) discuss various challenges with assessing student ability and using such assessments in research.

Second, unlike income or temperature, cognitive ability does not have any natural metric. Test scores are reported on different and arbitrary scales. Moreover, there is no reason to believe that the scores reported from standardized assessments have an interval property—that is, a one-unit change having the same meaning at every point on the scale. For example, it is unlikely that an increase from 400 to 450 on the Scholastic Aptitude Test (SAT) represents the same improvement in student knowledge as an increase from 700 to 750. Like utility, measured achievement is best thought of as ordinal, not cardinal. Bond and Lang (2013) illustrate how the unavoidably arbitrary nature of test scaling can influence empirical analysis. They show that the change in the Black-white test score gap between kindergarten and third grade can be as small as zero or as large as 0.6 standard deviations depending on the assumptions made about how to scale standardized assessment results.

Third, the use of standardized scores (subtracting the mean and dividing by the standard deviation) is not a magic bullet. As Jacob and Rothstein (2016) explain, standardized scores are no more comparable across tests or samples than raw or scale scores because standardization is relative to some norming population, which in practice can be small and nonrepresentative. Consider a common empirical result that interventions aimed at younger children tend to have larger effects on standardized test scores than do those aimed at older children. Cascio and Staiger (2012) point out that this pattern may be attributable to the fact that the variance in individual ability increases with age. Given that older children have been exposed to more out-of-school influences as well as more opportunities to learn (or not) in school, it is quite plausible that the true variance of ability increases with age. In this case, one would expect to see the pattern of declining effects with age even in the absence of any true relationship.

These issues complicate the analysis Hoxby proposes. For example, if test scores of young children have more measurement error than adult scores, the correlation between child and adult scores in a region could increase with the child's age even if the relationship between the underlying ability of children and adults remained the same. Even more broadly, I would argue that it is extremely difficult to determine how to measure advanced skills in common standardized assessments, much less create common measures across assessments targeted at different ages. With the limitations imposed by the available data—from the SAT, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), National Education Longitudinal Study (NELS), and Program for the International Assessment of Adult Competencies (PIAAC)—it is even harder to do so.

The approach Hoxby takes of using percentiles of the distribution is certainly reasonable, but far from ideal. Moreover, I would argue that the focus on advanced skills is not necessary. In practice, especially in the aggregate that Hoxby studies, the correlation between the average test scores and the proportion of individuals scoring advanced is quite high. The point is that cognitive skill is important and that the geographic distribution of such skill is increasingly unequal in the United States.

In sum, while there are compelling reasons to believe that skill is distributed more evenly (in terms of geography) among young children than adults, and that the association between child and adult skill increases during adolescence, I think this is a hard case to make statistically and I think the magnitude of the changes is entirely unclear.

THE EDUCATIONAL DECLINE IN ADOLESCENCE While I have some quibbles with how Hoxby uses cognitive assessments, I am willing to accept that adolescence is a critical period in a child's education. In addition to the neurological changes in the brain that children experience as adolescents, teenagers (at least in the United States) live in a unique culture that does not prioritize education. As the famous sociologist of education James Coleman wrote in his 1965 classic, *Adolescents and the Schools*, the "adolescent culture . . . shows little interest in education and focuses the attention of teen-agers on cars, dates, sports, popular music, and other matters just as unrelated to school" (72).

For years education researchers have recognized that children experience a relative drop in academic performance during adolescence. The US Department of Education administers an assessment to a nationally representative set of students in grades four, eight, and twelve every other year. The assessment, known as the National Assessment of Educational Progress (NAEP), is designed to consistently measure student learning across grades and over time. Moreover, a collection of educators and psychometricians work hard to link test scores to more meaningful measures of skill mastery. Based on this well-developed link between performance on test items and skill mastery, NAEP reports the percentage of students scoring in each of four categories: below basic, basic, proficient, and advanced. A consistent finding across subjects and years is the decline in student performance over the school years. On the 2019 NAEP mathematics assessment, for example, 19 percent of fourth graders scored below basic compared with 40 percent of twelfth graders. While 9 percent of students were designated advanced in fourth grade, only 3 percent demonstrated this level of mastery by twelfth grade. A similar pattern is apparent when comparing the United States to other countries. In a 2015 international assessment of fourth graders,

the United States ranked thirteenth out of forty-three countries. In a similar assessment given to 15-year-olds across the globe in 2018, the United States ranked twenty-ninth.[3]

POTENTIAL SOLUTIONS The evidence presented by Hoxby emphasizes the importance of reaching adolescents in skill deserts—typically poor, rural communities that lack a critical mass of highly skilled adults. What do we know about strategies to serve this population? Despite the chaos and poor quality that have characterized many children's experiences with online learning during the COVID-19 pandemic, educational technology offers some promise for helping boost achievement of children in disadvantaged, rural communities.

In discussing educational technology, it is important to distinguish between the use of virtual instruction as a supplement and as a substitute for face-to-face learning. A large body of research shows that student outcomes are substantially lower in online environments compared with traditional brick-and-mortar schooling. This is true at both the K-12 and the post-secondary level (Figlio, Rush, and Yin 2013; Hart, Friedmann, and Hill 2018; Heppen and others 2017; Bettinger and others 2015; Woodworth and others 2015).

However, research also points to several ways in which educational technology can enhance learning. First, there is evidence that technology can expand access to high-quality content and instruction. Students in under-resourced schools tend to have fewer advanced placement (AP) offerings, elective courses, and foreign language courses compared with their peers (Barker 1985). Similarly, high-poverty schools are also less likely to offer summer school, where students can retake a course they failed during the year (Watson and Gemin 2008). The best evidence on whether simply improving access to different courses through virtual schooling affects students' academic outcomes comes from a large-scale random assignment study carried out in Maine and Vermont (Heppen and others 2012). Sixty-eight schools that had not historically offered Algebra I to eighth graders were randomly assigned to either a treatment group, which was given access to an online Algebra I course, or a control group, which did not receive access. Algebra-ready students in treated schools showed

3. Data for fourth graders come from the 2015 Trends in International Mathematics and Science Study (TIMSS) and data for 15-year-olds come from the 2018 Programme for International Student Assessment (PISA). Assessment results can be accessed through the National Center for Education Statistics, https://nces.ed.gov/surveys/international/ide/. Forty-three countries reported scores on both of these exams and are thus used to generate the calculations reported here.

improvements on test scores and took more advanced courses in high school. Goodman, Melkers, and Pallais (2019) illustrate this potential at the postsecondary level. They study the Georgia Institute of Technology's online MS in computer science. Using a regression discontinuity design that exploits the admissions threshold, they show that the online option substantially increases overall enrollment.[4]

There is also compelling evidence that so-called intelligent tutoring systems, which provide instruction, practice, and feedback tailored to the needs of individual students, can improve student achievement. Some of the most compelling evidence comes from large randomized trials conducted in developing countries such as India (Banerjee and others 2007; Muralidharan, Singh, and Ganimian 2019). However, there is also evidence that intelligent tutoring is effective in the United States (Escueta and others 2020). As important, existing research suggests important lessons for developers and practitioners. For example, teachers in the United States often face challenges in effectively implementing computer-aided technology in a classroom setting (Drummond and others 2011; Pane and others 2010, 2013). In addition, experience to date suggests that computer-aided learning alone—in the absence of personal interaction between an adult and child—is not particularly effective, and the most effective programs are "blended," meaning they include some group-based instruction along with some individual student work with a personalized learning technology (Muralidharan, Singh, and Ganimian 2019).

The spread of online instruction driven by pandemic lockdowns spurred renewed interest in technology-aided education resources. The recently established National Student Support Accelerator provides a comprehensive set of resources for school districts and communities interested in implementing high-intensity, technology-supported tutoring programs.[5] Researchers and educators are taking a careful look at the potential of these strategies. If this work paves the way to reach at-risk adolescents in skill deserts, then it would truly be a silver lining of the pandemic.

4. There is even greater evidence in the developing economy context. For example, Bianchi, Lu, and Song (2020) find that the Chinese government's push to expand computer-assisted learning in rural communities substantially improved educational attainment.

5. Annenberg Institute for School Reform at Brown University, "National Student Support Accelerator," https://studentsupportaccelerator.com/.

REFERENCES FOR THE JACOB COMMENT

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122, no. 3: 1235–64.

Barker, Bruce. 1985. "Curricular Offerings in Small and Large High Schools: How Broad Is the Disparity?" *Research in Rural Education* 3, no. 1: 35–38.

Bettinger, Eric, Lindsay Fox, Susanna Loeb, and Eric Taylor. 2015. "Changing Distributions: How Online College Classes Alter Student and Professor Performance." Working Paper 15–10. Stanford, Calif.: Stanford Center for Education Policy Analysis.

Bianchi, Nicola, Yi Lu, and Hong Song. 2020. "The Effect of Computer-Assisted Learning on Students' Long-Term Development." Working Paper 28180. Cambridge, Mass.: National Bureau of Economic Research.

Bond, Timothy N., and Kevin Lang. 2013. "The Evolution of the Black-White Test Score Gap in Grades K–3: The Fragility of Results." *Review of Economics and Statistics* 95, no. 5: 1468–79.

Cascio, Elizabeth U., and Douglas O. Staiger. 2012. "Knowledge, Tests, and Fadeout in Educational Interventions." Working Paper 18038. Cambridge, Mass.: National Bureau of Economic Research.

Chetty, Raj, and Nathaniel Hendren. 2018a. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." *Quarterly Journal of Economics* 133, no. 3: 1107–62.

Chetty, Raj, and Nathaniel Hendren. 2018b. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *Quarterly Journal of Economics* 133, no. 3: 1163–228.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106, no. 4: 855–902.

Coleman, James S. 1965. *Adolescents and the Schools.* New York: Basic Books.

Drummond, Kathryn, Marjorie Chinen, Teresa Garcia Duncan, H. Miller, Lindsay Fryer, Courtney Zmach, and Katherine Culp. 2011. *Impact of the Thinking Reader [R] Software Program on Grade 6 Reading Vocabulary, Comprehension, Strategies, and Motivation: Final Report*. Washington: National Center for Education Evaluation and Regional Assistance.

Escueta, Maya, Andre Joshua Nickow, Philip Oreopoulos, and Vincent Quan. 2020. "Upgrading Education with Technology: Insights from Experimental Research." *Journal of Economic Literature* 58, no. 4: 897–996.

Figlio, David, Mark Rush, and Lu Yin. 2013. "Is It Live or Is It Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning." *Journal of Labor Economics* 31, no. 4: 763–84.

Friedman-Krauss, Allison H., W. Steven Barnett, Karin A. Garver, Katherine S. Hodges, G. G. Weisenfeld, and Beth Ann Gardiner. 2021. *The State of Preschool 2020*. New Brunswick, N.J.: National Institute for Early Education Research, Rutgers.

Goodman, Joshua, Julia Melkers, and Amanda Pallais. 2019. "Can Online Delivery Increase Access to Education?" *Journal of Labor Economics* 37, no. 1: 1–34.

Hart, Cassandra, Elizabeth Friedmann, and Michael Hill. 2018. "Online Course-Taking and Student Outcomes in California Community Colleges." *Education Finance and Policy* 13, no. 1: 42–71.

Heppen, Jessica B., Nicholas Sorensen, Elaine Allensworth, Kirk Walters, Jordan Rickles, Suzanne Stachel Taylor, and Valerie Michelman. 2017. "The Struggle to Pass Algebra: Online vs. Face-to-Face Credit Recovery for At-Risk Urban Students." *Journal of Research on Educational Effectiveness* 10, no. 2: 272–96.

Heppen, Jessica B., Kirk Walters, Margaret Clements, Ann-Marie Faria, Cheryl Tobey, Nicholas Sorensen, and Katherine Culp. 2012. *Access to Algebra I: The Effects of Online Mathematics for Grade 8 Students*. Washington: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, and US Department of Education.

Jacob, Brian, and Jesse Rothstein. 2016. "The Measurement of Student Ability in Modern Assessment Systems." *Journal of Economic Perspectives* 30, no. 3: 85–108.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review* 109, no. 4: 1426–60.

Pane, John F., B. A. Griffin, D. McCaffrey, R. Karam, L. Daugherty, and A. Phillips. 2013. "Does an Algebra Course with Tutoring Software Improve Student Learning?" Santa Monica, Calif.: RAND Corporation.

Pane, John F., Daniel F. McCaffrey, Mary Ellen Slaughter, Jennifer L. Steele, and Gina S. Ikemoto. 2010. "An Experiment to Evaluate the Efficacy of Cognitive Tutor Geometry." *Journal of Research on Educational Effectiveness* 3, no. 3: 254–81.

Watson, John, and Butch Gemin. 2008. *Using Online Learning for At-Risk Students and Credit Recovery*. Vienna, Va.: North American Council for Online Learning.

Woodworth, James L., M. E. Raymond, K. Chirbas, M. Gonzalez, Y. Negassi, W. Snow, and C. Van Donge. 2015. *Online Charter School Study 2015*. Stanford, Calif.: Center for Research on Educational Outcomes.

GENERAL DISCUSSION Carol Graham emphasized two points that were made in both Caroline Hoxby's presentation and Erik Hurst's discussion: adolescents are likely giving up if they are not on a trajectory toward college and low-skill jobs are disappearing. Graham noted that these patterns line up with the main findings from her work on deaths of despair. In her research, Graham finds that deaths of despair are high among non-college-educated white people and are highly correlated with low levels of hope. Graham observed that many of the areas with low cognitive skills that Hoxby highlights in her paper are predominantly white areas with low

levels of civic education, low levels of trust in science, and high rates of opioid use. Graham suggested that these trends can create a vicious feedback loop in which prime-age males experience low levels of participation in the labor force and high levels of despair.

Graham then argued that these patterns present a bleak outlook for future generations given that we do not see large movements out of these areas of despair and into areas of opportunity. However, Graham noted that one source of optimism is the declining gap in educational attainment by race, which Hurst noted in his discussion as well. Graham pointed to survey data which show that both Black and Hispanic people are more likely to believe in the value of a college education than low-income white people, even if it may be harder for them to get one. Finally, Graham concluded that hope plays a central role in helping adolescents overcome barriers to receiving a college degree and performing well in the labor market.

Richard V. Reeves brought up Melissa Kearney and Phillip Levine's research on teenage pregnancy rates across the country.[1] Reeves suggested this work could complement Hoxby's geographic analysis of cognitive skill attainment and lead to some insights.

Janice Eberly asked Hoxby to comment on gender differences in cognitive skill levels. Eberly referred to Hurst's mention of gender differences in educational attainment and wondered if the same patterns hold in Hoxby's data.

Taking a step back, Hoxby explained that the paper she presented is the third in a series of three papers looking at cognitive skill patterns in the United States, where the first paper focuses on the long-term effects of not making the transition to advanced cognitive skills in early adolescence and the second paper analyzes natural experiments in cognitive skill interventions. The first paper finds that failing to transition to advanced cognitive skills in early adolescence does have long-term consequences, and the second one finds that successful interventions are more productive if done during adolescence than if done before or afterward, leading Hoxby to conclude that adolescence is a particularly malleable period in cognitive skill development.

Addressing Eberly's comment on gender differences, Hoxby noted that male and female educational trajectories look quite different from one another. Up until grade three, the trajectories are mostly similar, but afterward

1. See, for example, Melissa S. Kearney and Phillip B. Levine, "Why Is the Teen Birth Rate in the United States So High and Why Does It Matter?" *Journal of Economic Perspectives* 26, no. 2 (2012): 141–63.

they diverge, Hoxby explained. This is due to differences in timing of cognitive brain development in males and females. Boys usually fall about a year behind in terms of cognitive development and that gap remains up until the end of high school. Hoxby argued that the cognitive development lag that males experience during this critical period has longer-term effects, such as lower college degree attainment levels for males.

Jim Stock expressed a concern that some of the correlation patterns in Hoxby's work may be misinterpreted as causal relationships. Specifically, Stock referred to the associations between areas with low cognitive skills and trust in scientists' views on climate change. Stock expressed the need for caution in interpreting these associations so that causal relationships are not inappropriately attributed to complex correlations. He then asked Hoxby to comment on how her work can be used to better understand the causal mechanisms that may underlie these correlations.

In response to Stock's comments, Hoxby reiterated that her paper is careful to distinguish between correlations and causal relationships. She claimed that there may be several, non-mutually exclusive causal mechanisms having an impact on cognitive skill development, such as the teen pregnancy rates that Reeves noted earlier. However, Hoxby highlighted two associations that she believes to be the important takeaways from her paper. The first is that there is a lot of movement in cognitive development that happens in early adolescence. In other words, one's cognitive trajectory is not determined by the third grade. Hoxby argued that economists of education are often too fatalistic about children's potential to develop advanced cognitive skills later in life even if they fall behind in early childhood. The fact that adolescents are highly influenced by their environments is a reason to be optimistic about potential cognitive skill interventions that target adolescents. The second point Hoxby underscored is that there is a strong association between the cognitive skills of children and the skills of the adults around them. Hoxby then clarified that this does not necessarily imply that adults' cognitive skills causally affect their children's cognitive skills development. However, she reiterated that if we do want to get closer to understanding the underlying causal mechanisms, we need to carefully investigate the existing strong correlations.

Erik Hurst pondered whether the mechanisms that cause skill deserts within cities differ from the mechanisms that lead to skill deserts across cities.

In response to Hurst's question, Hoxby suggested that the same mechanisms may be at play in a large city as the ones occurring across cities. She used Chicago as an example of a city in which the differences in school

quality and other environmental factors across neighborhoods is similar to the differences found across cities in the country. To accentuate her point, Hoxby discussed the wide variation in middle school quality across the country. She argued that public middle schools are among the most neglected schools with the most teacher vacancies, leading to large differences in quality from one neighborhood to the next.