PUBLISHED FOR THE BROOKINGS INSTITUTION BY
JOHNS HOPKINS
UNIVERSITY PRESS

Brookings Papers
ON ECONOMIC ACTIVITY

BPEA
SPRING
2023

BROOKINGS

Janice Eberly and James H. Stock, Editors

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 2023

ABRAHAM and RENDELL
Where Are the Missing Workers? Anticipated and Unanticipated Labor Supply Changes in the Pandemic's Aftermath

BISTLINE, MEHROTRA, and WOLFRAM
Economic Implications of the Climate Provisions of the Inflation Reduction Act

PANEL: ACHARYA, CARVALHO, NECHIO, and PRASAD
Emerging Market Challenges

HONG and LUCAS
COVID-19 Credit Policies around the World: Size, Scope, Costs, and Consequences

GOLDBERG and REED
Is the Global Economy Deglobalizing? If So, Why? And What Is Next?

FINKELSTEIN, McQUILLAN, ZIDAR, and ZWICK
The Health Wedge and Labor Market Inequality

Brookings Papers

ON ECONOMIC ACTIVITY

SPRING 2023

JANICE EBERLY
JAMES H. STOCK
Editors

PUBLISHED FOR THE BROOKINGS INSTITUTION BY
JOHNS HOPKINS
UNIVERSITY PRESS

Copyright © 2024 by
THE BROOKINGS INSTITUTION
1775 Massachusetts Avenue, N.W., Washington, D.C. 20036

ISSN 0007-2303
E-ISSN 1533-4465

Brookings Papers on Economic Activity is published twice a year in Spring and Fall by Johns Hopkins University Press, 2715 N. Charles Street, Baltimore, MD 21218-4363, USA. POSTMASTER: Send address changes to *Brookings Papers on Economic Activity*, Johns Hopkins University Press, Journals Division, P.O. Box 19966, Baltimore, MD 21211-0966.

All rights reserved. No portion of *Brookings Papers on Economic Activity* may be reproduced by any process or technique without the formal consent of the publisher. For more information, please visit the Press's permissions department at www.press.jhu.edu/cgi-bin/permissions.cgi.

The paper in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper). ♾

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 2023

KATHARINE G. ABRAHAM and LEA E. RENDELL
Where Are the Missing Workers? Anticipated and Unanticipated Labor Supply Changes in the Pandemic's Aftermath 1
Comment by Stephanie Aaronson 49
Comment by Yongseok Shin 63
General Discussion 70

JOHN E. T. BISTLINE, NEIL R. MEHROTRA, and CATHERINE WOLFRAM
Economic Implications of the Climate Provisions of the Inflation Reduction Act 77
Comment by Jason Furman 158
Comment by Kenneth Gillingham 167
General Discussion 178

PANEL on EMERGING MARKET CHALLENGES

VIRAL V. ACHARYA
India at 75: Replete with Contradictions, Brimming with Opportunities, Saddled with Challenges 185

CARLOS CARVALHO and FERNANDA NECHIO
Challenges to Disinflation: The Brazilian Experience 217

ESWAR PRASAD
Has China's Growth Gone from Miracle to Malady? 243

Comment by Kenneth Rogoff 271
General Discussion 282

GEE HEE HONG and DEBORAH LUCAS
COVID-19 Credit Policies around the World: Size, Scope, Costs, and Consequences 289
Comment by Alan J. Auerbach 334
General Discussion 341

PINELOPI K. GOLDBERG and TRISTAN REED
Is the Global Economy Deglobalizing? If So, Why? And What Is Next? 347
Comment by Pol Antràs 397
Comment by Douglas A. Irwin 410
General Discussion 417

AMY FINKELSTEIN, CASEY McQUILLAN, OWEN ZIDAR, and ERIC ZWICK
The Health Wedge and Labor Market Inequality 425
Comment by David Cutler 476
Comment by Angus Deaton 487
General Discussion 499

PURPOSE

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. Working drafts of the papers are presented and discussed at conferences held twice each year, and the final versions of the papers and comments along with summaries of the general discussions 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; General Motors Company; the National Science Foundation, under grant no. 2048708; the Smith Richardson Foundation; and State Farm Mutual Automobile Insurance Company. We express appreciation to Brevan Howard Reserve Services for supporting *BPEA*'s mission and activities. We gratefully acknowledge Dina Axelrad Perry for establishing the George L. Perry and William C. Brainard *BPEA* Chair.

The views expressed by the authors, discussants, and conference participants in *BPEA* are strictly those of the authors, discussants, and conference participants and not of the Brookings Institution. As an independent think tank, the Brookings Institution does not take institutional positions on any issue. Conference drafts and recordings of the Spring 2023 *BPEA* Conference can be accessed at https://www.brookings.edu/events/bpea-spring-2023-conference/.

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 print or e-subscriptions or single copies for institutions or individuals, please visit www.press.jhu.edu/journals or contact Johns Hopkins University Press, Journals Division, P.O. Box 19966, Baltimore, MD 21211-0966 (USA). Phone: 410-516-6987 • Toll free: 1-800-548-1784 (US/Canada only) • Email: jrnlcirc@jh.edu. Archived issues of *BPEA* are also available on Project MUSE (https://muse.jhu.edu/journal/52) and JSTOR (www.jstor.org).

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

Henry Aaron *Brookings Institution*
Stephanie Aaronson *Federal Reserve Board of Governors*
John Abowd *US Census Bureau*
Katharine Abraham *University of Maryland*
Viral Acharya *New York University Stern School of Business*
George Akerlof *Georgetown University*
Pol Antràs *Harvard University*
Alan Auerbach *University of California, Berkeley*
Laurence Ball *Johns Hopkins University*
Jose Maria Barrero *Instituto Tecnológico Autónomo de México*
Joe Beaulieu *Brevan Howard*
Ben Bernanke *Brookings Institution*
Carola Binder *Haverford College*
John Bistline *Electric Power Research Institute*
Olivier Blanchard *Peterson Institute for International Economics*
Alan Blinder *Princeton University*
William Brainard *Yale University*
Steven Braun *Council of Economic Advisers*
Jesse Bricker *Federal Reserve Board of Governors*
Markus Brunnermeier *Princeton University*
Ralph Bryant *Brookings Institution*
Gary Burtless *Brookings Institution*
David Byrne *Federal Reserve Board of Governors*
Carlos Carvalho *Kapitalo Investimentos & PUC-Rio*
Anne Case *Princeton University*
Ajay Chhibber *George Washington University*
Gabriel Chodorow-Reich *Harvard University*
Constance Citro *Committee on National Statistics at National Academies*
Gerald Cohen *University of North Carolina Kenan Institute of Private Enterprise*
David Cutler *Harvard University*
Steven Davis *Hoover Institution*
Angus Deaton *Princeton University*
Ryan Decker *Federal Reserve Board of Governors*
Mark Doms *Congressional Budget Office*
John Driscoll *Federal Reserve Board of Governors*
Janice Eberly *Northwestern University*
Wendy Edelberg *Brookings Institution*
Bruce Fallick *Federal Reserve Bank of Cleveland*
Ana Fernandes *World Bank*
James Feyrer *Dartmouth College*
Matthew Fiedler *Brookings Institution*
Amy Finkelstein *Massachusetts Institute of Technology*

Kristin Forbes *MIT Sloan School of Management*
Rebecca Freeman *Bank of England*
Benjamin Friedman *Harvard University*
Jason Furman *Harvard University*
Joseph Gagnon *Peterson Institute for International Economics*
Amy Ganz *Aspen Economic Strategy Group*
Tarek Ghani *Washington University in St. Louis*
Kenneth Gillingham *Yale University*
Pinelopi Goldberg *Yale University*
Robert Gordon *Northwestern University*
Yuriy Gorodnichenko *University of California, Berkeley*
Joshua Gotbaum *Brookings Institution*
Francois Gourio *Federal Reserve Bank of Chicago*
Carol Graham *Brookings Institution*
Adam Guren *Boston University*
Joe Gyourko *University of Pennsylvania Wharton School*
Robert Hall *Stanford University*
John Haltiwanger *University of Maryland*
James Hamilton *University of California, San Diego*
Peter Blair Henry *Stanford University*
Brad Hershbein *W. E. Upjohn Institute for Employment Research*
Douglas Holtz-Eakin *American Action Forum*
Gee Hee Hong *International Monetary Fund*
Caroline Hoxby *Stanford University*
Ethan Ilzetzki *London School of Economics*
Douglas Irwin *Dartmouth College*
Şebnem Kalemli-Özcan *University of Maryland*
Melissa Kearney *University of Maryland*
Amit Khandelwal *Yale University*
Michael Kiley *Federal Reserve Board of Governors*
Aaron Klein *Brookings Institution*
Jeffrey Kling *Congressional Budget Office*
Donald Kohn *Brookings Institution*
Arvind Krishnamurthy *Stanford University*
Randy Kroszner *University of Chicago Booth School of Business*
Raphael Lee *French Department of Treasury*
Adam Looney *University of Utah*
Deborah Lucas *Massachusetts Institute of Technology*
Maria Luengo-Prado *Federal Reserve Bank of Boston*
Greg Mankiw *Harvard University*
Robert McClelland *Urban-Brookings Tax Policy Center*
Casey McQuillan *Princeton University*
Neil Mehrotra *Federal Reserve Bank of Minneapolis*
Bruce Meyer *University of Chicago*
Laurence Meyer *Monetary Policy Analytics dba LH Meyer*
Gian Maria Milesi-Ferretti *Brookings Institution*
Jeffrey Miron *Harvard University*
Frederic Mishkin *Columbia University*
Robert Moffitt *Johns Hopkins University*
Adele Morris *Federal Reserve Board of Governors*
Fernanda Nechio *Federal Reserve Bank of San Francisco*
Christopher Nekarda *Federal Reserve Board of Governors*
Maurice Obstfeld *University of California, Berkeley*
Sanjay Patnaik *Brookings Institution*
George Perry *Brookings Institution*
Jonathan Pingle *UBS*

David Popp *Syracuse University*
Richard Portes *London Business School*
Eswar Prasad *Brookings Institution & Cornell University*
Brian Prest *Resources for the Future*
Sarah Reber *Brookings Institution*
Alessandro Rebucci *Johns Hopkins University*
Tristan Reed *World Bank*
David Reifschneider *formerly of the Federal Reserve Board of Governors*
Carmen Reinhart *Harvard University*
Lea Rendell *University of Maryland*
Giovanni Ricco *University of Warwick*
Kenneth Rogoff *Harvard University*
Liliana Rojas-Suarez *Center for Global Development*
David Romer *University of California, Berkeley*
Glenn Rudebusch *Brookings Institution*
John Sabelhaus *Brookings Institution*
Jake Schild *Bureau of Labor Statistics*
Louise Sheiner *Brookings Institution*
Yongseok Shin *Washington University in St. Louis*
Tara Sinclair *US Department of Treasury*
Mark Steinmeyer *Smith Richardson Foundation*
Jón Steinsson *University of California, Berkeley*
James Stock *Harvard University*
Michael Strain *American Enterprise Institute*
Martin Stuermer *International Monetary Fund*
Phillip Swagel *Congressional Budget Office*
Angelos Theodorakopoulos *Aston Business School*
Anne Thompson *Massachusetts Institute of Technology*
Stan Veuger *American Enterprise Institute*
Masazumi Wakatabe *Waseda University*
William Wascher *Federal Reserve Board of Governors*
Beatrice Weder di Mauro *Centre for Economic Policy Research*
David Wessel *Brookings Institution*
Jon Willis *Federal Reserve Bank of Atlanta*
Justin Wolfers *University of Michigan*
Catherine Wolfram *Massachusetts Institute of Technology*
Susan Woodward *Sand Hill Econometrics*
Jonathan Wright *Johns Hopkins University*
Owen Zidar *Princeton University*
Eric Zwick *University of Chicago Booth School of Business*

Elijah Asdourian *Brookings Institution*
Maricarmen Barron Esper *Brookings Institution*
Haowen Chen *Brookings Institution*
Alexandria Cordero *Brookings Institution*
Siobhan Drummond *Brookings Institution*
Simran Kalkat *Brookings Institution*
Isabel Leigh *Brookings Institution*
Conrad Milhaupt *Brookings Institution*
Adam Sedlak *Brookings Institution*
David Skidmore *Brookings Institution*

KATHARINE G. ABRAHAM
University of Maryland

LEA E. RENDELL
University of Maryland

Where Are the Missing Workers? Anticipated and Unanticipated Labor Supply Changes in the Pandemic's Aftermath

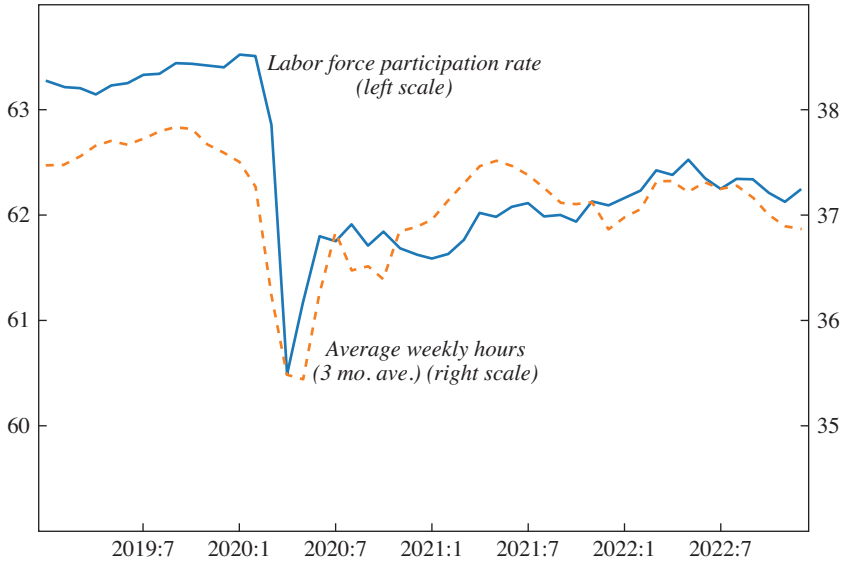
ABSTRACT Labor force participation and average hours of work both fell sharply at the beginning of the COVID-19 pandemic. Neither had fully recovered by the end of 2022. The drop in participation between December 2019 and December 2022 implies a loss of 3 million people from the labor force; the decline in average hours over the same period translates to the equivalent of 2.6 million fewer workers. Demographic and other trend factors that predated the pandemic explain most of the participation shortfall. Taken together, COVID-19-related health effects and the persistent (though shrinking) effects of the fear of contracting COVID-19 more than explain the rest. In contrast, pre-pandemic factors account for little of the shortfall in hours. COVID-19-related health effects account for perhaps 40 percent of that decline, but we are unable to explain the majority of the hours shortfall. We speculate that the lower level of hours in the post-pandemic period may reflect a shift in the desired balance between work and other aspects of workers' lives.

The US labor force participation rate has been trending downward since the beginning of the 21st century (Abraham and Kearney 2020), falling on average about 0.2 percentage point per year. At the start of the pandemic, however, participation plummeted, dropping more than 3 percentage points on a seasonally adjusted basis over just two months. As can

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. The authors are not currently an officer, director, or board member of any organization with a financial or political interest in this paper.

Brookings Papers on Economic Activity, Spring 2023: 1–48 © 2024 The Brookings Institution.

Figure 1. Labor Force Participation (Monthly) and Average Weekly Hours (Three-Month Moving Average), Seasonally Adjusted, January 2019 to December 2022



Sources: Authors' tabulations of Current Population Survey data.

Note: Series constructed using monthly data downloaded from the IPUMS data set (Flood and others 2022) and reweighted to wedge effects of new population controls introduced in January 2022 and January 2023 backward. Individuals with a job but not at work were assigned zero hours. Monthly estimates seasonally adjusted using X-12 seasonal adjustment command in Eviews, applying multiplicative X-11 method and auto X-12 seasonal and trend filter.

be seen in figure 1, about half of that initial drop was quickly reversed, but as of December 2022, the seasonally adjusted participation rate was still more than a percentage point below its December 2019 level. Average weekly hours of work as measured in the Current Population Survey (CPS) also fell sharply at the onset of the pandemic and, as of late 2022, had not fully recovered. Seasonally adjusted average weekly hours for the three months centered on December 2022 were about 0.7 hour below their level for the three months centered on December 2019.[1]

1. We count people who had a job but did not work during the survey reference week as having zero hours. Even after seasonal adjustment, the weekly hours series is noisy. To make the trends in the data more easily visible, figure 1 plots a centered three-month average rather than the monthly numbers. As explained later in the paper, all of our analysis makes use of data adjusted to wedge the effects of the new population controls introduced in the CPS in January 2022 and January 2023 back over an appropriate prior period rather than treating them as changes that occurred in a single month. This means that our estimates do not exactly match published CPS estimates.

In this paper, we explore more systematically what happened to the labor force participation rate and average weekly hours of work over the three years ending in December 2022.[2] Ongoing declines in participation due to the aging of the baby boomers have long been anticipated. Although rising educational attainment has partially offset the effects of population aging, taken together these demographic changes accounted for a bit more than a 0.5 percentage point decline in participation between December 2019 and December 2022, leaving an unexplained decline of about 0.6 percentage point. Even prior to the pandemic, moreover, participation among younger and prime-age adults generally had been trending downward. Although increasing participation among older adults has partially offset these declines, extrapolating the pre-pandemic within-group trends in addition to accounting for changes in demographic mix can explain an additional 0.3 percentage point decline in the participation rate from December 2019 through December 2022. This leaves a decline of just 0.3 percentage point that is not explained by preexisting factors.

The post-pandemic decline in average hours per week reduced total hours worked almost as much as the decline in participation. In contrast to the decline in participation, however, demographic changes and pre-pandemic trends explain almost none of this drop. Our paper highlights the importance of hours for understanding how labor supply changed over the post-pandemic period.

Measured relative to changes projected based on a continuation of pre-existing within-group trends, the post-pandemic shortfalls in participation and hours were not uniform across age and education groups. Participation among adults age 25–54 dropped more than anticipated for those without a college degree but changed little among the college-educated. Among adults age 65 and older, in contrast, there were *larger* unanticipated participation rate declines among the college-educated than among those with less education, though in this case both groups' participation rates fell significantly relative to their previous trends. Participation net of demographic and trend influences fell for white non-Hispanic individuals, but changed little for Black non-Hispanic and Hispanic

2. Although excess mortality and reductions in immigration associated with the pandemic also have affected the size of the labor force through their effects on the size of the population (Powell 2022; Hobijn and Şahin 2022), we focus here on what has happened to the willingness of the resident population to supply their labor to the market.

individuals. The unanticipated declines in participation were a bit larger for women than for men, but the reverse is true for the unanticipated declines in hours.

Various hypotheses have been offered for why, as late as December 2022, participation and hours had not yet fully rebounded to the levels that would have been predicted prior to the pandemic. We explore (and in some cases extend) available evidence related to several widely cited explanations: that improvements in households' balance sheets attributable to federal safety net spending or to the rising stock market and increases in house prices had a dampening effect on labor supply; that COVID-19 itself led to lower participation and shortened hours of work; and that fear of COVID-19 slowed the return of workers to the labor market.

We conclude that, although they contributed to improvements in household balance sheets, neither federal transfers to households during the pandemic nor rising asset prices contributed in any substantial way to lower overall labor supply. Rising house prices, however, could have made it possible for older homeowners adversely affected by other shocks to retire earlier than they otherwise might have done.

Available data suggest that missed work time due to COVID-19 infections explains a portion of the shortfall in hours we document. Some prior studies have estimated very large effects of long COVID—lingering health problems consequent to a COVID-19 infection—on labor supply. Although we do not find the largest previously estimated effects to be plausible, we conclude that long COVID explains much if not all of the modest shortfall in participation relative to what one might have anticipated prior to the pandemic and a portion of the shortfall in hours. By the end of 2022, fear of COVID-19 had become less important as a reason for nonparticipation but continued to play a role. Insofar as working even a shortened in-person work schedule would have involved potential COVID-19 exposure, however, fear of COVID-19 seems unlikely to have contributed significantly to the shortfall in hours.

Taken together, demographic changes, preexisting trends, COVID-19 health effects, and continuing fear of COVID-19 (more than) explain the post-pandemic shortfall in participation we document. Even after considering all of these factors, however, much of the unanticipated decline in average weekly hours remains unexplained. This leads us to speculate that a reevaluation of the balance between work and other activities may be a part of the explanation for the reduction in hours we document. Whether this change will be permanent remains to be seen.

I. How Did Labor Supply Change in the Post-Pandemic Labor Market?

The central question we seek to answer is why, despite the strength in the labor market, neither participation nor weekly hours of work had fully recovered to pre-pandemic levels nearly three years later. We look first at the raw changes in participation and hours between December 2019 and December 2022, then examine series from which we net out the effects of demographic changes and, for a second counterfactual, also net out the effects of preexisting trends. These estimates imply that much of the shortfall in participation compared to pre-pandemic levels should have been anticipated based on demographic and other trend factors. In contrast, almost none of the shortfall in weekly hours of work can be explained by demographic changes or pre-pandemic group-specific trends.

I.A. Developing a Benchmark for Assessing Post-Pandemic Participation and Hours

Prior to the start of the pandemic, the labor market had been experiencing a long-running cyclical recovery, with unemployment having reached lows not seen since the late 1960s. Had the labor market been less strong at the end of 2022 than in late 2019 or early 2020, an explanation for the post-pandemic shortfalls in participation and hours might have been that they were due to weaker labor demand conditions. In fact, however, available data suggest that the labor market was at least as tight at the end of 2022 as it had been three years earlier. The seasonally adjusted unemployment rate stood at 3.6 percent in December 2019 and was 3.5 percent in December 2022. If anything, the elevated levels of both job vacancies and quits prevailing at the end of 2022 suggest a tighter labor market than implied by the unemployment rate (Domash and Summers 2022). This leads us to interpret the shortfalls we observe as a product of changes in labor supply, not labor demand. In the remainder of the paper, we compare participation and hours as of December 2022 to their levels in December 2019. One caution about the interpretation of our results is that labor supply typically responds to tightening labor markets with a lag. Although this process had been largely completed by the end of 2022 (Hobijn and Şahin 2022), data from the months following the endpoint for our analysis suggest it had not fully played itself out.

Even absent the pandemic, changing population demographics over the past three years, most especially the aging of the population and ongoing

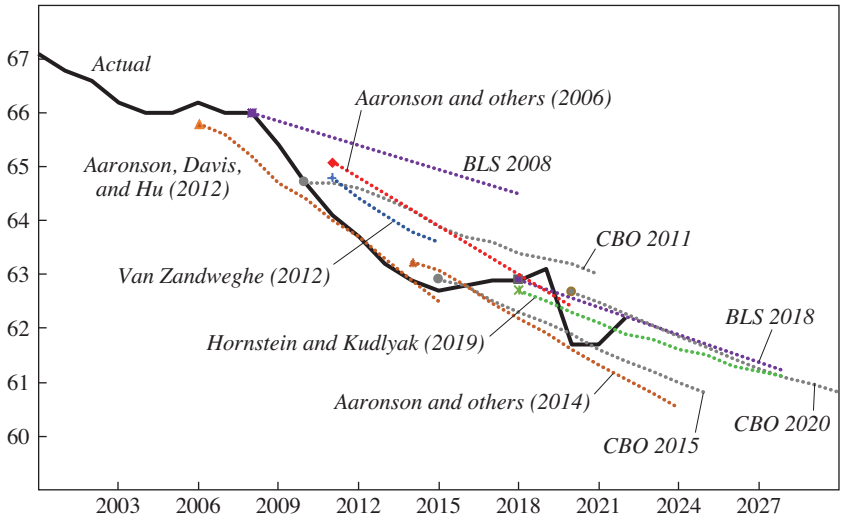
increases in educational attainment at given ages, would have affected participation and possibly hours. Because participation drops off significantly at older ages, it has long been anticipated that the aging of the baby boomers would put downward pressure on the overall participation rate. Rising education levels have worked in the opposite direction, as once their schooling is completed more-educated individuals have higher participation rates than those with lower levels of education, but have only partially offset the effects of aging. The net effect of these demographic changes on participation has been negative (Aaronson, Davis, and Hu 2012; Aaronson and others 2006; Aaronson and others 2014; Montes 2018; Hornstein and Kudlyak 2019). Average weekly hours vary less with age and education than participation, but also should be adjusted for these same demographic factors.

Some discussions of post-pandemic labor force participation implicitly have set a benchmark that incorporates a continuation of subgroup participation trends during the years immediately preceding the pandemic. As unemployment fell and the labor market tightened during this period, within-group participation rates generally were rising. Extrapolating those short-term trends generates a benchmark against which the number of "missing workers" is very large. Given the significant cyclical upswing in participation that already had occurred by the beginning of 2020, however, continuation of the immediate pre-pandemic participation trends is an unrealistic counterfactual.[3]

Although it would not be appropriate to extrapolate the short-term pre-pandemic trends and treat that as the benchmark for recent experience, there have been notable longer-term trends in participation that might well have been expected to continue. Participation among young adults (16–24 years old) and prime-age adults (25–54 years old) generally has fallen since about 2000; in contrast, participation among older adults (55 years old and older) generally has risen (Abraham and Kearney 2020). Analysts have modeled these trends in different ways. The ten-year labor force projections developed by the Bureau of Labor Statistics, for example, essentially extrapolate preexisting participation trends for specific subgroups (US Bureau of Labor Statistics 2022b). Others have directly estimated participation rate models for demographic subgroups that contain cyclical, structural, and cohort variables, though different analysts have made differing choices about exactly what variables to include in the model,

3. Hobijn and Şahin (2022) make a similar point.

Figure 2. Actual and Projected Cyclically Neutral Labor Force Participation Rates, 2000–2028



Source: Hobijn and Şahin (2022), adapted with authors' permission.

Note: Actual figures are published CPS estimates; BLS projections available at https://www.bls.gov/emp/; and CBO forecasts available at https://www.cbo.gov/data/budget-economic-data. Other estimates from Aaronson, Davis, and Hu (2012); Aaronson and others (2006); Aaronson and others (2014); Hornstein and Kudlyak (2019); and Van Zandweghe (2012). Markers show vintage of projection or forecast.

how the cohort effects are specified, and other model features (Aaronson, Davis, and Hu 2012; Aaronson and others 2006; Aaronson and others 2014; Montes 2018; Hornstein and Kudlyak 2019). Despite the differences in approach—and some notable differences in what the models imply about the level of the cyclically neutral labor force participation rate—there is broad agreement about the combined effects of demographic changes and other factors on the overall trend in labor force participation.

This can be seen in figure 2, adapted from a recent paper by Hobijn and Şahin (2022).[4] Both across models and across different vintages of the same model, there are significant differences in the level of the cyclically neutral participation rate. In every case, however, the models imply that the overall participation rate has been trending downward by between

4. We thank Bart Hobijn and Ayşegül Şahin for sharing the data underlying their chart with us. Our chart includes a slightly different set of previous projections than were included in their chart but otherwise is similar to theirs.

0.2 and 0.3 percentage point per year. The adjustments we make to account for changes in demographic mix and, in some calculations, preexisting within-group trends, described below, are consistent with these projections.

I.B. Adjusting Participation and Hours for Demographics and Preexisting Group-Specific Trends

We measure participation and hours using data from the CPS; the underlying microdata were downloaded from the IPUMS database (Flood and others 2022). As in the series displayed in figure 1, our average hours calculations assign zero hours to people who had a job but did not work during the survey reference week. One complication in working with these data is that the population controls used to determine the weighting for CPS estimates are updated each January but the weights and associated estimates for previous months are not adjusted to reflect this updated information. The January 2021 changes to the population controls had a negligible effect on key estimates (US Bureau of Labor Statistics 2021), but both the January 2022 and the January 2023 changes raised the estimated labor force participation rate. The new controls introduced in January 2022 reflected information from the 2020 Census showing that the population was younger than had previously been believed. Because younger adults are more likely than older adults to be working or seeking work, using the new controls would have raised the December 2021 labor force participation rate by 0.3 percentage point (Montes, Smith, and Dajon 2022; US Bureau of Labor Statistics 2022a). This discrepancy should be thought of as having accumulated over the period since population controls incorporating information from the 2010 Census were introduced rather than occurring all at once. An implication is that the participation rate just prior to the start of the pandemic was a bit higher and the size of the post-pandemic decline in participation through December 2021 a bit larger than suggested by published statistics.[5] The new controls introduced in January 2023 produced smaller changes, but their use would have raised the December 2022 labor force participation rate by 0.1 percentage point (US Bureau of Labor Statistics 2023). Incorporating the January 2023 controls thus has the effect of shrinking the post-pandemic decline in participation, partially reversing the effects of incorporating the January 2022 controls.

5. Montes, Smith, and Dajon (2022) make a similar point in the context of discussing the growth in retirees as a share of the population through October 2022, which they argue would have been understated had they not taken the January 2022 change in population controls into account.

To measure the changes in participation and hours more accurately, we use weight adjustment factors developed by Bauer and others (2023) that wedge the changes captured by the new population controls introduced in January 2022 and January 2023 backward to the beginning of the period over which each of them is likely to have accumulated.

Given estimates of the declines in participation and hours between December 2019 and December 2022 incorporating the weight adjustments just described, we first would like to know how much of the changes we see can be attributed to factors whose effects were anticipatable prior to the onset of the pandemic. Subtracting the influence of those anticipatable factors should tell us how much remains to be explained by factors related to the pandemic and its aftermath. We begin by removing the effects of changes in demographic mix from the data and then, additionally, adjust for the potential effects of preexisting within-group participation rate trends.

Disaggregating across any mutually exclusive set of population subgroups, the change in the labor force participation rate between any two periods can be written as:

$$\Delta\left(LFPR\right)_{t_0,t_1} = \sum_i s_{i,t_0}\Delta\left(LFPR\right)_{i,t_0,t_1} + \sum_i \left(LFPR\right)_{i,t_0}\Delta s_{i,t_0,t_1} + \sum_i \Delta s_{i,t_0,t_1}\Delta\left(LFPR\right)_{i,t_0,t_1} \tag{1}$$

where *LFPR* is the estimated labor force participation rate, *s* is a specific group's share of the overall population, *i* indexes the different mutually exclusive groups, and t_0 and t_1 are the start and end of the time period over which the change is measured. We can express the change in average hours similarly, but with employment shares in place of population shares. The first term in these equations captures the effects of within-group changes in participation or hours, holding population or employment shares fixed. A series constructed based on that term tells us how much participation or hours has changed net of the effects of changes in demographic mix. The second term captures the effect of changes in group shares, holding each group's initial participation rate or average hours constant. The third term summarizes any potential interaction effects. Although these can matter in some contexts, they are consistently negligible for the decompositions we report.

The decomposition just described treats each group's participation rate or hours in the initial period as the relevant benchmark for evaluating subsequent within-group changes. Since persistent trends reflect factors that

might have a continuing influence, it also may be appropriate to net out the effects of any preexisting within-group trends. We can modify equation (1) to do this:

$$\Delta\left(LFPR\right)_{t_0,t_1} = \sum_i s_{i,t_0}\left[\Delta\left(LFPR\right)_{i,t_0,t_1} - TREND_{i,t_0,t_1}\right] + \sum_i\left[\left(LFPR\right)_{i,t_0}\Delta s_{i,t_0,t_1} + s_{i,t_0}TREND_{i,t_0,t_1}\right] + \sum_i \Delta s_{i,t_0,t_1}\Delta\left(LFPR\right)_{i,t_0,t_1} \tag{2}$$

In equation (2), *TREND* is the estimated preexisting long-term within-group trend for a group extrapolated over the time period of interest. The first term in the equation is now the unanticipated change in participation after netting out the effects of both demographic changes and preexisting within-group trends. The second term is the change that should have been anticipated based on the combination of those factors. The third term is unchanged and, as before, represents potential interaction effects. A similar decomposition can be carried out for the change in hours using employment shares rather than population shares.

We begin by constructing series that fix the age by education composition of the population or employment, as appropriate, but allow within-age-and-education-group participation and hours to change as they did. For these series, we hold constant the shares in each of thirteen detailed age groups (16–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, and 75 and older). Because those age 16–24 are still in the process of completing their education and the decision whether to remain in school is itself an endogenous outcome, we do not hold educational attainment constant within the 16–19 or 20–24 age groups. For those age 25 and older, we fix the shares within each detailed age group in each of the following four educational attainment categories—less than a high school education, high school or the equivalent, some college, or a bachelor's degree or higher.[6]

6. Among those age 16–24, participation rates and hours of work are markedly higher for those who are out of school than for those who are still enrolled. Because the shares of 16- to 19-year-olds and 20- to 24-year-olds who are out of school has risen, holding constant the shares who are in versus out of school would leave slightly more of the change in participation to be explained by within-group changes as opposed to changes in population demographics.

Starting with the age-by-education-adjusted series, we also construct series that additionally net out the effects of within-group trends from the first term of equation (2) or its hours analogue. For this purpose, the preexisting trend in each of the detailed age groups (for those under age 25) or age by education groups (for those age 25 and older) is estimated using annual data for the period from 2000 through 2019.

Figure 3 shows the result of netting out the effects of demographic changes and preexisting group-specific trends. For participation, the figure displays seasonally adjusted monthly numbers. The hours series shown in the figure are three-month moving averages.

As shown in figure 3, panel A, the seasonally adjusted labor force participation rate we estimate fell by 1.2 percentage points between December 2019 and December 2022. On an age-and-education-adjusted basis, participation was about 0.6 percentage point lower in December 2022 than it had been in December 2019. Allowing for preexisting within-age-and-education-group trends, participation was about 0.3 percentage point lower in December 2022. This second calculation suggests that most of the post-pandemic decline in participation may be attributable to factors that predated the pandemic, though there was still a modest shortfall as of December 2022.

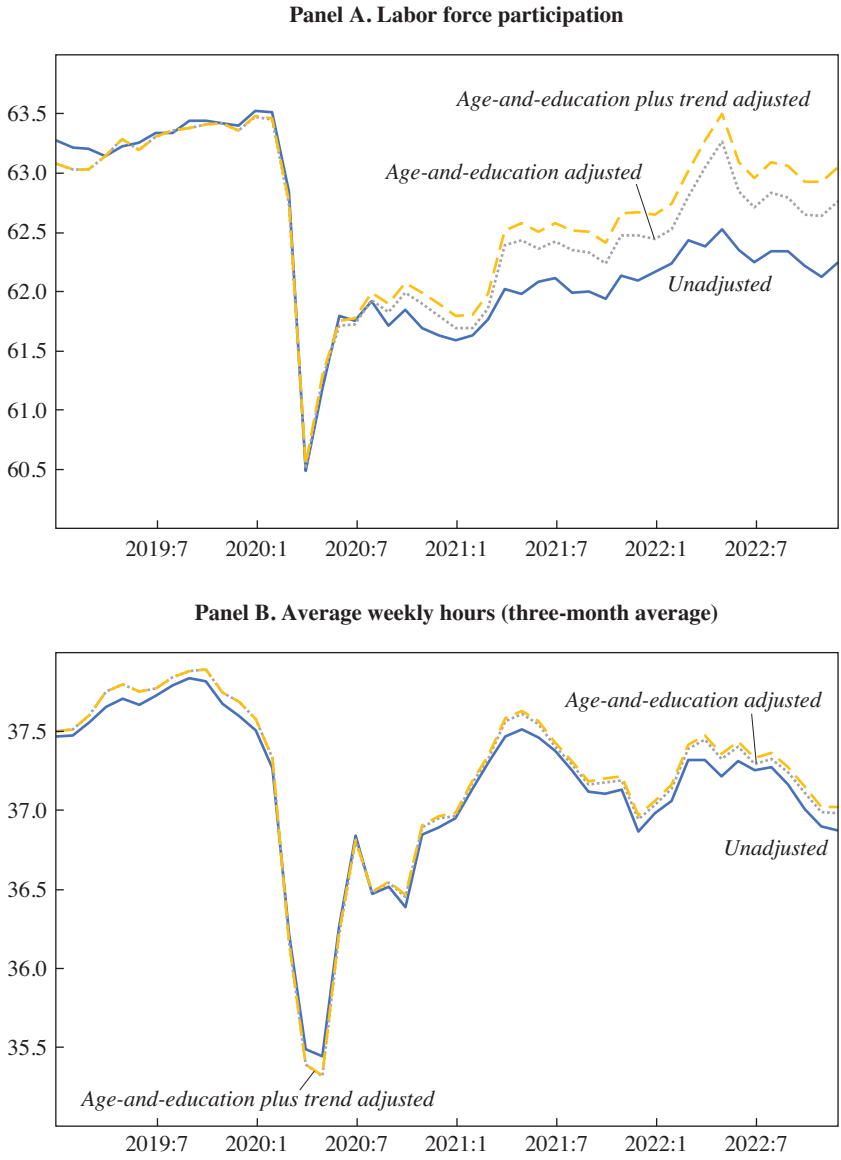
Panel B shows corresponding series for hours of work. Average weekly seasonally adjusted hours measured as a three-month moving average were 36.9 hours in December 2022 compared to 37.6 hours in December 2019 or 0.7 hour lower. Although there are differences in average weekly hours across some of the employment subgroups, adjusting for the age-by-education composition of employment does not greatly affect the measured overall net change in average hours. Removing the effects of long-term preexisting within-group trends in average hours also has little effect on the measured overall net change.

To be slightly more formal about the relative importance of changes along different margins to the change in the aggregate supply of hours, consider the following identity:

$$\frac{TOTHRS_t}{POP_t} = \frac{LF_t}{POP_t} * \frac{EMP_t}{LF_t} * AVEHRS_t \tag{3}$$

where *TOTHRS* is aggregate hours worked per week; *POP* is the population age 16 and older; *LF* is the number of people age 16 and older in the labor force; *EMP* is the number of employed people age 16 and older; and *AVEHRS* is the average weekly hours worked by those who are employed.

Figure 3. Unadjusted and Adjusted Labor Force Participation and Average Weekly Hours, January 2019 to December 2022



Sources: Authors' tabulations of Current Population Survey data.

Note: Series constructed using monthly data downloaded from the IPUMS data set (Flood and others 2022) and reweighted to wedge effects of new population controls introduced in January 2022 and January 2023 backward. Individuals with a job but not at work were assigned zero hours. Monthly estimates seasonally adjusted using X-12 seasonal adjustment command in Eviews, applying multiplicative X-11 method and auto X-12 seasonal and trend filter. For adjusted series, component series seasonally adjusted and then summed.

In this identity, $\frac{LF}{POP}$ is just the labor force participation rate (*LFPR*) and $\frac{EMP}{LF}$ can be expressed as one minus the unemployment rate (1 – *UR*). Substituting into equation (3), taking the natural log and then differencing gives us:

$$(4) \qquad \Delta \ln\left(TOTHRS\right) - \Delta \ln\left(POP\right) = \Delta \ln\left(LFPR\right) + \Delta \ln\left(1 - UR\right) + \Delta \ln\left(AVEHRS\right)$$

This expression decomposes the change in aggregate weekly hours per adult member of the population into pieces attributable to changes in the labor force participation rate, the employment rate (one minus the unemployment rate), and average weekly hours among the employed. Carrying out the equation (4) decomposition for December 2022 compared to December 2019 (and using monthly data that are not seasonally adjusted for all of the terms), the natural logarithm of aggregate hours per person age 16 and older fell by 0.0328, roughly a 3.2 percent decline. Of this decline, about 55 percent was due to declining labor force participation (which fell about 1.8 percent) and about 48 percent to the decline in average weekly hours (which fell about 1.6 percent), with a small offsetting positive residual attributable to the 0.1 percentage point decrease in the unemployment rate between the two endpoints.[7]

A simple way to put these numbers into perspective is to ask how many missing labor force participants the changes along each of the three margins represent. The approximately 1.8 percent decline in labor force participation we observe between December 2019 and December 2022 reduced the size of the labor force by about 3.0 million in December 2022. Based on our separate decomposition of the factors that have contributed to the participation decline, roughly 1.5 million of this shortfall remains after allowing for changes in the age-education composition of the population and a bit under 0.8 million remains after allowing, in addition, for the continuation of preexisting within-group participation trends. The approximately

7. Lee, Park, and Shin (2023) report results from a similar analysis that decomposes the change in aggregate hours from 2019 through 2022 into a piece attributable to changes in the employment to population ratio and changes in average weekly hours. They find that changes in average hours have been roughly as important as changes in the employment to population ratio in explaining the post-pandemic aggregate hours shortfall.

1.6 percent decline in average weekly hours translates into the equivalent of a 2.6 million reduction in the size of the labor force. In contrast to the participation figures, very little of this is attributable to demographic changes or preexisting trends. Holding everything else constant, the change in the unemployment rate would have *added* a small amount to the available workforce (about 0.1 million people).

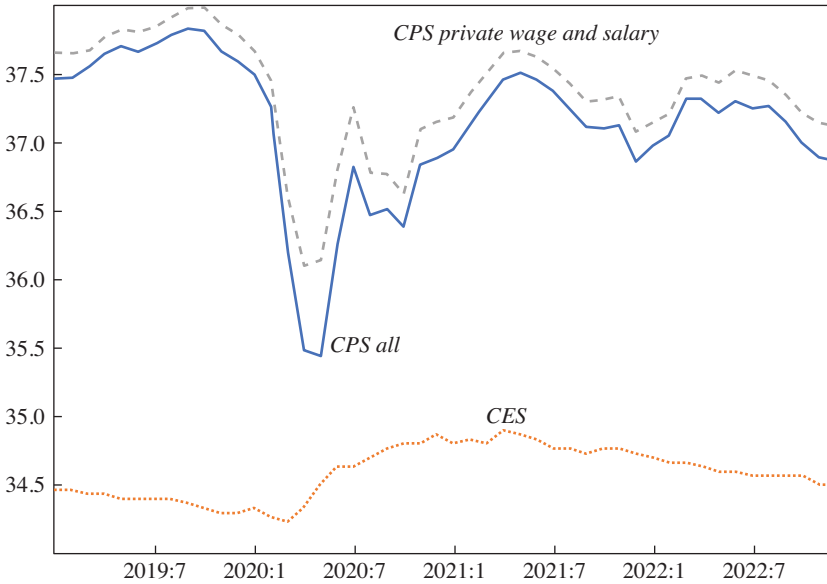
I.C. An Aside on the Measurement of Hours

In these decompositions, net of the effects of demographic changes and preexisting within-group trends, the shortfall in average weekly hours is a more important part of the post-pandemic labor supply shortfall than the corresponding shortfall in labor force participation. Hours as estimated in the Current Employment Statistics (CES) survey (the monthly payroll survey), however, give a very different sense of how hours moved during the post-pandemic period. Figure 4 shows three-month moving averages of seasonally adjusted estimates for both CPS and payroll survey hours for the period from January 2019 through December 2022. As already discussed, weekly hours of work per employed person as measured in the CPS fell sharply at the start of the pandemic and had not fully recovered by the end of 2022. Weekly CPS hours for people whose main job is a private sector wage and salary job—a group that aligns more closely with the payroll survey universe—behaved similarly to overall weekly CPS hours. On a seasonally adjusted basis, the private sector wage and salary series declined from about 37.7 hours in December 2019 to about 37.1 hours in December 2022. In contrast, seasonally adjusted weekly paid hours per private sector job as measured in the payroll survey *rose* following the onset of the pandemic, increasing from 34.3 hours in December 2019 to 35.0 hours in January 2021 before eventually falling back to 34.4 hours as of December 2022, 0.1 hour above their starting level.

Although we cannot fully reconcile the different movements in the CPS and payroll survey hours, we can identify two factors that likely have been at work. First, the CPS measures hours *worked*, whereas the payroll survey measures hours *paid*. Second, whereas the CPS is a measure of hours per *person*, the CES survey is a measure of hours per *job*. Both help to explain why the two hours series behaved differently over the post-pandemic period.

The gap between hours worked and hours paid is especially important for salaried workers. Although salaried workers' actual weekly hours may vary considerably, they are paid for a fixed number of hours per week. We

Figure 4. Average Weekly Hours, Current Population Survey and Current Employment Statistics Survey, January 2019 to December 2022



Sources: Authors' tabulations of Current Population Survey data and Bureau of Labor Statistics.

Note: CPS hours data downloaded from the IPUMS data set (Flood and others 2022) and reweighted to wedge effects of new population controls introduced in January 2022 and January 2023 backward. In CPS data, individuals with a job but not at work were assigned zero hours. Monthly estimates seasonally adjusted using X-12 seasonal adjustment command in Eviews, applying multiplicative X-11 method and auto X-12 seasonal and trend filter. CES survey hours data seasonally adjusted series as published by Bureau of Labor Statistics. All series centered three-month moving averages.

are able to distinguish salaried workers from hourly workers in the quarter of the CPS sample that belongs to the outgoing rotation groups (those in the survey sample for either the fourth or the eighth of eight interviews). Seasonally adjusted average weekly hours worked for private sector wage and salary workers in this outgoing rotation group sample were about 0.9 hour lower in the three months centered on December 2022 than in the three months centered on December 2019. To gauge what this decline might have looked like had hours paid rather than hours worked been recorded for the salaried workforce, we recalculated the numbers with the hours of full-time salaried workers fixed at 40 hours per week and those of part-time salaried workers fixed at 20 hours per week. The resulting series, which should more closely approximate an hours paid series, exhibits a net decline of just over half as much as the original series.

This back-of-the-envelope calculation does not fully capture how converting the measurement from hours worked to hours paid would affect the post-pandemic CPS estimates, as it does not account for the potential effects of changes in paid time off among the hourly workforce. During the pandemic, reflecting the impact of COVID-19, there were increases both in the share of workers absent from work for health-related reasons and in the share working part-time rather than full-time for health-related reasons.[8] The increases in these shares have been larger among private sector hourly workers than for the wage and salary workforce as a whole. Suppose that affected hourly workers otherwise would have worked the same hours as in the previous month, provided they were working rather than absent (for the first group) or working full-time rather than part-time (for the second group). Under these assumptions, the extra work time missed by hourly workers due to absences or cutting back to part-time in the three months centered on December 2022 compared to the three months centered on a typical December would have reduced overall average weekly hours worked by about 0.2 hour. To the extent that these workers had access to paid sick leave, hours paid would not reflect this reduction in hours.[9]

The other major conceptual difference between the CPS and payroll hours series is that the former measures hours per person while the latter measures hours per job. This difference means that changes over time in the prevalence of multiple job holding may cause the series to diverge. Data from the CPS indicate that the seasonally adjusted multiple job holding rate among workers whose primary job was as a private sector employee fell from 4.7 percent in December 2019 to a low of 3.8 percent in April 2020, stayed low through the end of 2020, and then recovered, slowly and unevenly, to 4.7 percent in December 2022. All else the same, the initial decline in multiple job holding would have led hours per employee as measured in the CPS to fall. In addition, because second jobs generally involve fewer hours per week than the average job, a decline in multiple job holding can be expected to have caused average hours per job, the concept measured in the payroll survey, to rise. Back-of-the-envelope calculations using CPS data suggest that, holding average hours per week

8. Goda and Soltas (2023) show that absences for health-related reasons among the workforce as a whole have been correlated with local area COVID-19 case counts and have grown more for workers in occupations that are more likely to require physical presence on the job, corroboration for the view that the increase can be attributed to COVID-19.

9. We use three-month averages for the calculations just described because information to distinguish salaried from hourly workers is available only for the outgoing rotation groups and the sample sizes are correspondingly smaller.

on the primary and secondary jobs held by employees whose primary job was in the private sector at their December 2019 level, the reduction in the multiple job holding rate through April 2020 would have reduced average weekly hours per private sector employee by about 0.1 hour. It also would have raised average weekly hours per job they held by about 0.2 hour.[10] Multiple job holding thus helps to explain why, at the pandemic's start, CPS hours and payroll survey hours initially moved in opposite directions.

In sum, the differences between measuring hours worked versus hours paid and hours per person versus hours per job appear to explain much if not all of the divergence in the behavior of CPS and payroll survey hours since the pandemic's start. Our interest lies with understanding changes in labor supply over this period and, for that purpose, the CPS numbers are the conceptually more appropriate measure. In what follows, we make use of hours worked as measured in the CPS.

I.D. Whose Labor Supply Fell?

The estimates we have presented thus far suggest that there were notable changes in aggregate labor supply during the post-pandemic period relative to the changes that should have been expected based on shifts in population demographics or the combination of shifts in population demographics and preexisting trends. This naturally raises the question of which groups in the population were responsible for these changes. Table 1 summarizes the results of an accounting exercise that decomposes the overall change in the participation rate between December 2019 and December 2022 into the components shown in either equation (1) or equation (2). These calculations make use of non-seasonally adjusted data reweighted as discussed above to more appropriately distribute the effects of the new population controls introduced in January 2022 and January 2023. The mutually exclusive groups used for the calculations are defined using detailed age groups and, for those age 25 and older, completed education. The decomposition shown in panel A accounts only for changes in age and education mix; panel B additionally incorporates the projected effects of a continuation of within-group trends. In each case, we are interested primarily in the row in the table that shows the decline in participation in

10. A limitation of this calculation is that some of the second jobs held by people whose main job is as a private sector employee are in self-employment. It also does not count second jobs as employees held by people who are primarily self-employed. Abraham and others (2023) report that, in the CPS, about 75 percent of second jobs held by people who are employees on their main job are also employee jobs, as are about 40 percent of second jobs held by the much smaller share of people who are self-employed on their main job.

Table 1. Decomposition of Percentage Point Change in Labor Force Participation Rate, without and with Trend Adjustments, December 2019 to December 2022

	Contribution to overall change in LFPR					
	Overall	*Men*	*Women*	*White non-Hispanic*	*Black non-Hispanic*	*Hispanic*
Total change	−1.13	−1.06	−1.25	−1.51	−0.54	−0.71
Panel A: Decomposition without accounting for within-group trends						
Within-age-and-education group participation changes	−0.59	−0.59	−0.62	−0.78	−0.31	−0.30
16–24, all	0.05	0.04	0.05	0.10	−0.15	0.01
25–54, less than college	−0.39	−0.41	−0.37	−0.41	−0.06	−0.54
25–54, college plus	−0.01	−0.02	−0.04	−0.03	−0.17	0.18
55–64, less than college	0.03	−0.02	0.09	0.01	−0.05	0.11
55–64, college plus	0.00	0.02	−0.03	−0.06	0.27	−0.01
65 and older, less than college	−0.11	−0.12	−0.12	−0.14	−0.15	−0.14
65 and older, college plus	−0.16	−0.10	−0.20	−0.25	0.00	0.08
Age-and-education share changes	−0.55	−0.46	−0.64	−0.74	−0.15	−0.47
Interactions	0.01	−0.01	0.01	0.02	−0.07	0.06
Panel B: Decomposition accounting for within-group trends						
Within-age-and-education group participation changes less trend effects	−0.30	−0.19	−0.43	−0.58	0.11	−0.09
16–24, all	0.31	0.37	0.26	0.34	−0.02	0.33
25–54, less than college	−0.18	−0.18	−0.13	−0.20	0.20	−0.43
25–54, college plus	0.02	0.03	−0.07	−0.03	−0.09	0.16
55–64, less than college	−0.01	−0.05	0.05	−0.03	0.00	0.01
55–64, college plus	−0.04	−0.02	−0.08	−0.11	0.24	−0.03
65 and older, less than college	−0.18	−0.18	−0.19	−0.22	−0.20	−0.19
65 and older, college plus	−0.22	−0.16	−0.27	−0.33	−0.03	0.07
Age-and-education share changes plus trend effects	−0.84	−0.86	−0.83	−0.95	−0.57	−0.69
Interactions	0.01	−0.01	0.01	0.02	−0.07	0.06

Panel C: Change in participation rate (average across detailed component groups)						
16–24, all	0.32	0.29	0.35	0.80	–0.90	0.07
25–54, less than college	–1.32	–1.28	–1.37	–1.67	–0.17	–1.19
25–54, college plus	–0.04	–0.09	–0.18	–0.17	–1.02	1.48
55–64, less than college	0.22	–0.18	0.62	0.29	0.80	–0.73
55–64, college plus	0.00	0.37	–0.59	–0.88	7.54	–0.56
65 and older, less than college	–0.82	–0.98	–0.80	–0.87	–1.25	–1.62
65 and older, college plus	–2.49	–1.43	–3.20	–3.02	0.10	4.58
Panel D: Change in participation rate relative to preexisting trend (average across detailed component groups)						
16–24, all	2.12	2.41	1.83	2.73	–0.15	1.62
25–54, less than college	–0.60	–0.57	–0.49	–0.81	0.56	–0.95
25–54, college plus	0.08	0.18	–0.34	–0.16	–0.54	1.35
55–64, less than college	–0.13	–0.49	0.32	–0.06	0.88	–1.42
55–64, college plus	–0.69	–0.44	–1.55	–1.63	6.89	–1.23
65 and older, less than college	–1.32	–1.47	–1.26	–1.36	–1.68	–2.20
65 and older, college plus	–3.40	–2.33	–4.33	–3.91	–0.76	3.80

Sources: Authors' analysis of Current Population Survey data.

Note: Data adjusted to smooth changes in estimates due to introduction of new population controls in January 2022 and January 2023. Preexisting within-group trends estimated over 2000–2019 period. Detailed age groups are 16–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, and 75 plus. Detailed education categories for those age 25 and older are less than high school, high school, some college, and bachelor's degree or higher (labeled as college plus). Race and ethnicity categories are mutually exclusive. Race and ethnicity breakouts exclude non-Hispanic individuals who identify as Asian; American Indian, Aleut, or Eskimo; Hawaiian or Pacific Islander; or more than one race. People in these excluded groups represent less than 10 percent of the population.

excess of what we would have anticipated based on these factors. Estimates for the whole population are reported in the table's first column; the remaining columns show results by gender and by race and ethnicity. The table also shows the contribution of different broad age by education groups to the overall decline in participation relative to expectations and, for reference, the average change in the labor force participation rate within each of these same broad groups, again measured relative to expectations. The top row in table 1 shows that, not seasonally adjusted, the overall participation rate declined by about 1.1 percentage points between December 2019 and December 2022, but also that there were some differences across groups.[11] Participation declined noticeably more for white non-Hispanic individuals than for either Black non-Hispanic individuals (a 1.0 percentage point difference) or Hispanic individuals (a 0.8 percentage point difference). Women's participation rate declined slightly more than men's.

Given previously expected demographic changes and preexisting within-group participation rate trends, much of the overall change in participation could have been anticipated prior to the pandemic. After adjusting for changes in demographic mix, overall participation was roughly 0.6 percentage point lower than expected as of December 2022; also allowing for the effects of preexisting trends, it was only about 0.3 percentage point lower.

Focusing on the estimates that account for preexisting trends, overall participation declined most sharply for adults age 65 and older, and especially so for those with a college degree. The declines in participation for adults age 55–64, while less pronounced, also were larger for those with a college degree than for those with less education. These patterns are consistent with the finding by Montes, Smith, and Dajon (2022) that excess retirements increased more during the pandemic for those with a college education than for those without a college degree. Among adults age 25–54, in contrast, only the participation of those with less than a college degree fell relative to the pre-pandemic trend. These declines were partially offset by unanticipated within-group participation *increases* (relative to trend) among those age 16–24.

There are also differences in the patterns by gender and by race and ethnicity, though because of smaller sample sizes for some of the subgroups used in the decompositions, these results should be interpreted with

11. The declines in seasonally adjusted participation and hours plotted in figure 1 and figure 3 differ slightly from the non-seasonally adjusted declines reported in table 1 and table 2. Because the large and unusual movements in these series at the start of the pandemic make it more difficult to identify the appropriate seasonal factors, we chose to use non-seasonally adjusted data for the decompositions.

caution. The declines in participation for college-educated adult women age 55–64 and age 65 and older were larger relative to what would have been projected prior to the pandemic than those for men. Among college-educated adults age 65 and older, where the adjusted overall participation rate dropped substantially, the decline relative to trend was much smaller for Black non-Hispanic individuals, and participation relative to trend actually rose for highly educated older Hispanic individuals. The overall average decline in participation net of previous trends among everyone age 65 and older (not shown) was larger for white non-Hispanic individuals (2.2 percentage points) than for either Black non-Hispanic (1.5 percentage points) or Hispanic individuals (1.2 percentage points). Consistent with these findings, Montes, Smith, and Dajon (2022) report that excess retirements increased more during the pandemic for older white adults than for older Black or Hispanic adults.

For completeness, we include a similar decomposition of the post-pandemic change in average weekly hours. As can be seen in table 2, changes in the age by education composition of the employed population do not account for any appreciable share of the overall change in average weekly hours in the post-pandemic period—essentially all the decline reflects within-group hours reductions—nor do preexisting within-group trends in hours explain much of the observed change. Men's average weekly hours fell more than women's. Hours fell slightly less for Hispanic workers than for white non-Hispanic workers and actually increased slightly for Black non-Hispanic workers.

II. How Did the Pandemic Affect Participation and Hours?

There are several channels through which the pandemic—and the public policy response to the pandemic—could have contributed to the post-pandemic shortfall in participation and hours just documented. One potential explanation is that improvements in household balance sheets due to federal financial assistance during the pandemic may have slowed the return to work (Ferguson 2023). A related explanation is that increasing household wealth, the rising stock market, and especially, growing house prices associated with the pandemic-induced increase in the demand for housing in many locations could have led some people to exit the labor market (Coile 2022; Favilukis and Li 2023). Contracting COVID-19 leads directly to missed work hours. In addition, long COVID symptoms may have forced some affected individuals to withdraw from the labor force or cut back more permanently on the time they spend at work (Bach 2022a,

Table 2. Decomposition of Change in Average Weekly Hours, without and with Trend Adjustments, December 2019 to December 2022

	Contribution to overall change in average weekly hours					
	Overall	*Men*	*Women*	*White non-Hispanic*	*Black non-Hispanic*	*Hispanic*
Total change	−0.59	−0.88	−0.29	−0.73	0.14	−0.78
Panel A: Decomposition without accounting for within-group trends						
Within-age-and-education group hours changes	−0.58	−0.89	−0.27	−0.68	0.06	−0.79
16–24, all	−0.04	−0.07	0.00	−0.08	0.16	0.05
25–54, less than college	−0.35	−0.46	−0.24	−0.36	−0.08	−0.69
25–54, college plus	−0.10	−0.20	−0.01	−0.10	−0.09	−0.09
55–64, less than college	−0.06	−0.10	0.00	−0.08	−0.08	0.00
55–64, college plus	−0.01	−0.02	−0.01	−0.01	0.13	−0.04
65 and older, less than college	0.00	−0.01	0.01	−0.01	0.02	0.01
65 and older, college plus	−0.03	−0.03	−0.03	−0.03	0.00	−0.04
Age-education share changes	−0.03	−0.01	−0.04	−0.07	−0.03	0.00
Interactions	0.02	0.02	0.02	0.02	0.11	0.01
Panel B: Decomposition accounting for within-group trends						
Within-age-and-education group hours changes less trend effects	−0.54	−0.74	−0.37	−0.70	0.10	−0.59
16–24, all	0.00	−0.03	0.02	−0.06	0.19	0.14
25–54, less than college	−0.33	−0.39	−0.24	−0.36	−0.04	−0.62
25–54, college plus	−0.06	−0.12	−0.05	−0.07	−0.07	−0.05
55–64, less than college	−0.08	−0.11	−0.02	−0.11	−0.10	0.00
55–64, college plus	−0.02	−0.03	−0.03	−0.02	0.12	−0.03
65 and older, less than college	−0.02	−0.03	−0.01	−0.03	0.00	0.01
65 and older, college plus	−0.04	−0.04	−0.04	−0.05	−0.01	−0.04
Age-and-education share changes plus trend effects	−0.07	−0.16	0.06	−0.05	−0.08	−0.20
Interactions	0.02	0.02	0.02	0.02	0.11	0.01

Panel C: Change in average weekly hours (average across detailed component groups)						
16–24, all	−0.30	−0.60	−0.03	−0.75	1.21	0.33
25–54, less than college	−0.96	−1.18	−0.71	−1.16	−0.17	−1.29
25–54, college plus	−0.35	−0.80	−0.02	−0.35	−0.37	−0.54
55–64, less than college	−0.19	−0.27	−0.02	−0.41	−0.03	−0.03
55–64, college plus	−0.22	−0.37	−0.17	−0.17	3.34	−1.35
65 and older, less than college	0.13	−0.16	0.39	−0.20	0.72	0.49
65 and older, college plus	−0.87	−0.79	−1.04	−0.89	−0.20	−5.28
Panel D: Change in average weekly hours relative to preexisting trend (average across detailed component groups)						
16–24, all	−0.01	−0.29	0.20	−0.52	1.47	0.89
25–54, less than college	−0.90	−0.99	−0.72	−1.15	−0.08	−1.16
25–54, college plus	−0.22	−0.48	−0.17	−0.24	−0.31	−0.32
55–64, less than college	−0.33	−0.34	−0.18	−0.58	−0.13	−0.07
55–64, college plus	−0.25	−0.41	−0.41	−0.24	3.15	−1.06
65 and older, less than college	−0.45	−0.70	−0.18	−0.78	0.11	0.47
65 and older, college plus	−1.28	−1.26	−1.51	−1.30	−0.61	−5.43

Source: Authors' analysis of Current Population Survey data.

Note: Data adjusted to smooth changes in estimates due to introduction of new population controls in January 2022 and January 2023. Preexisting within-group trends estimated over 2000–2019 period. Detailed age groups are 16–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, and 75 plus. Detailed education categories for those age 25 and older are less than high school, high school, some college, and bachelor's degree or higher (labeled as college plus). Race and ethnicity categories are mutually exclusive. Race and ethnicity breakouts exclude non-Hispanic individuals who identify as Asian; American Indian, Aleut, or Eskimo; Hawaiian or Pacific Islander; or more than one race. People in these excluded groups represent less than 10 percent of the population.

2022b; Goda and Soltas 2023; Sheiner and Salwati 2022). Finally, continued fear of contracting COVID-19 could have kept people out of the labor force (Barrero, Bloom, and Davis 2022, 2023). We consider the evidence pertaining to these various explanations and then summarize the conclusions we draw from the available evidence.

II.A. Social Safety Net Expansions during 2020 and 2021

A common explanation for shortfalls in labor supply early in the pandemic was that relatively generous social safety net supports were discouraging people from looking for work (Morath and Chen 2020; Mitchell, Weber, and Cambon 2021). By the beginning of 2022, these added supports had come to an end, though household balance sheets remained healthier through the end of 2022 than might otherwise have been predicted, partially as a result of the earlier supports (Barnes and others 2022). The question is whether improvements in households' financial situations plausibly can explain the persistence of the shortfall in labor supply relative to prepandemic levels. To estimate how much the federal dollars that flowed to households during the pandemic might have affected labor supply, we first ask how large an increase in household wealth these payments represented. Then, we appeal to the literature that has estimated the elasticity of labor supply with respect to a pure increase in wealth to translate the increase in household wealth into a labor supply effect.

Estimating how much federal pandemic spending contributed to household wealth is tricky. Absent the pandemic, unemployment would not have spiked as it did and, for many households, unemployment benefits only partially replaced lost earnings. As a very generous upper bound estimate, we count as infusions to household balance sheets all of the approximately $441 billion the federal government spent on Federal Pandemic Unemployment Compensation, which added first $600 per week and later $300 per week to recipients' normal unemployment insurance (UI) payments; the approximately $85 billion spent on Pandemic Emergency Unemployment Compensation, which extended benefit durations; and the approximately $130 billion spent on Pandemic Unemployment Assistance, which made benefits available to the self-employed and others who ordinarily would not have qualified.[12] When the American Rescue Plan Act was passed in

12. These estimates come from United States Department of Labor, "Unemployment Insurance: Families First Coronavirus Response Act and Coronavirus Aid, Relief, and Economic Security (CARES) Act Funding to States," https://oui.doleta.gov/unemploy/docs/cares_act_funding_state.html, and include spending through October 8, 2022. We have subtracted spending that flowed to Puerto Rico and the Virgin Islands.

March 2021, the Congressional Budget Office estimated that the act's changes to the child tax credit would add about $85 billion to federal outlays during fiscal years 2021 and 2022 (CBO 2021). Spending on the three rounds of economic impact payments totaled about $850 billion (Parker and others 2022). Pandemic spending directed to households thus totaled roughly $1.6 trillion. As of the beginning of 2021, according to Census Bureau figures, there were about 250 million people age 18 and older living in the United States. On a per adult basis, federal pandemic spending directed to households thus amounted to an average of about $6,400.

In contrast to the voluminous literature on the effects of wages on labor supply, the literature on the labor supply effects of an increase in wealth is limited. A challenge is that increases in household wealth are seldom exogenous with respect to labor supply decisions. The best evidence we are aware of comes from studies of how lottery winnings affect the winners' subsequent labor supply. If people are myopic, the effects of an unexpected financial windfall such as winning a lottery could be larger in the short run than over time; if households are forward-looking, however, one would expect the effects to be smaller but more persistent. Consistent with the idea that households are forward-looking, Cesarini and others (2017) find that lottery winners spend their money slowly and that the labor supply effects of a lottery win are relatively stable for at least five years after the win. According to their elasticity estimates, winning a lottery valued at about $140,000 in 2010 dollars, equivalent to about $175,000 as of the middle of 2021, reduced winners' labor force participation rate by about 2 percentage points and their weekly hours by about 1.3 hours. Although one might expect these effects to be larger for older adults, they do not vary significantly by age group, at least not through age 65. Winning a lottery also affects a spouse's labor supply but only by about half as much as it affects the winner's. Using the elasticity estimates reported by Cesarini and others (2017), a wealth increase on the order of $6,400 would lower labor force participation by less than 0.1 percent and hours by less than 0.1 hour per week even after allowing for effects on spouses' labor supply. Given that much of the money the federal government directed toward households during the pandemic offset lost earnings rather than representing a net addition to household balance sheets, our assessment is that any negative effects of this spending on labor supply most likely were negligible.

II.B. Other Influences on Household Balance Sheets

Although discussions of how strong household balance sheets might be affecting labor supply often emphasize the role of federal payments during

the pandemic, these payments have not been the only or even the most important influence on households' financial well-being (Barnes and others 2022). While a significant number of households suffered serious economic harm due to pandemic shutdowns, others benefited initially from a rising stock market and rising home prices, the latter arguably a result of changes in the demand for housing induced by the pandemic. Stock prices later dropped and the housing market also cooled, though house prices remained significantly elevated. The group of people one might expect to be most affected by a run-up in asset prices are those close to retirement, for whom an increase in household wealth could make earlier retirement possible.

Previous research generally has concluded that short-term fluctuations in the stock market do not significantly affect the timing of retirement (Coile and Levine 2011; Goda, Shoven, and Slavov 2011). Research on the effects of short-term house price movements largely has come to a similar conclusion. Bosworth and Burtless (2010), for example, find no evidence in CPS data that trailing three-year house price increases in a respondent's state have a significant effect on the labor force participation of individuals age 55 to 74. Coile and Levine (2011) obtain similar results. Using data from the Health and Retirement Study, Farnham and Sevak (2016) find that year-over-year changes in housing prices do not have a significant effect on retirement rates.

Much of this research has been based on data for periods when the most significant variation in house prices was the large drop experienced in many markets during the Great Recession. The experience with rising house prices during the pandemic could have been different. Coile (2022), however, finds no evidence that state-level housing price changes during the pandemic were associated with increased labor market exits among those age 55–74. In contrast, in a recent paper, Favilukis and Li (2023), using American Community Survey (ACS) household data and Freddie Mac housing returns data, find that metropolitan area house price increases from 2016 through the middle of 2021 were associated with significantly lower labor force participation among older homeowners. For older renters and younger adults, however, house price increases were associated with *higher* participation; middle-aged adults' participation did not vary with house prices. Taken at face value, these results imply a negligible effect of the run-up in house prices on overall participation, but Favilukis and Li (2023) view it as explaining much if not all of the decline in participation among adults age 65 and older.

It is unclear, however, exactly what to make of these findings. The participation rate models estimated by Favilukis and Li (2023) include a variety of individual and metropolitan area controls in addition to their main house price change variable, but other factors still could explain the associations they find between house price increases and participation. For example, purchases by retirees of homes in smaller cities during the pandemic could have pushed up house prices in those areas while mechanically being associated with a lower participation rate for older adults. Further, Favilukis and Li are interested mainly in the role of rising house prices during the pandemic as a driver of lower participation among older homeowners, but their baseline model measures house price changes over a longer period. They do not obtain statistically significant results for house price increases that occurred after the start of 2020 while also controlling for pre-pandemic house price run-ups. Even setting these issues aside, there is a question about how any changes in older homeowners' participation behavior in areas where house prices were rising should be interpreted. Rather than house price changes driving declines in participation among older homeowners, it could be that something else—long COVID, fear of COVID-19, or simply the loss of a job—pushed older homeowners out of work. For those that had benefited from rising house prices, the fact that they had more home equity then made it possible for them not to return. Regardless, given that more-educated adults also are more likely to be homeowners, these findings are qualitatively consistent with the fact that, as documented earlier, the unanticipated declines in participation among college graduates age 65 and older have been larger than those for less-educated older adults.

II.C. COVID-19 Infections

Individuals who become infected with COVID-19 typically must take time off from work. Although active COVID-19 infections are unlikely to have an immediate effect on employment, they can be expected to reduce hours worked during the infection period. To estimate these effects, we use information from the CPS on the percentage of workers who were absent from work during the survey reference week for health-related reasons and on the percentage who worked part-time instead of their usual full-time for the same reasons. Both of these were higher in the post-COVID-19 period than they were on average from 2010 through 2019. We estimate lost hours for workers absent from their job for health-related reasons based on hours in the prior month for workers in this group who were working in the previous month. Similarly, we estimate lost hours for those

working part-time rather than full-time for health-related reasons by comparing current hours to hours in the prior month for workers in this group who were working full-time in the previous month. Our estimates suggest that, if the share of workers absent for health-related reasons and the share of workers who are normally full-time but temporarily part-time for health-related reasons had been equal in December 2022 to their pre-pandemic average for the month of December, average weekly hours for the workforce as a whole would have been between 0.1 and 0.2 hour higher. Our point estimate of the reduction in hours translates into the equivalent of roughly 710,000 workers.

II.D. Long COVID

Even after their initial recovery, a significant share of COVID-19 sufferers continue to experience debilitating symptoms that may include difficulty thinking or concentrating ("brain fog"), headaches, sleep problems, and depression or anxiety, among other symptoms (Aiyegbusi and others 2021). It would not be surprising if some of those suffering from long COVID have chosen to withdraw from the labor force or cut back on their hours.

A number of researchers have investigated the effects of long COVID on labor supply. One general approach to estimating the effect of long COVID on participation is to combine estimates of the number of people experiencing long COVID symptoms with estimates of the effect of experiencing long COVID on work activity. This is a sensible strategy, but unless the long COVID population is defined in the same way for estimating the number of long COVID sufferers as for estimating long COVID's labor supply impact, these estimates could be misleading. More specifically, a problem arises if the size of the long COVID population is defined in an inclusive fashion but the share of workers with long COVID symptoms who have withdrawn from the labor force or reduced their work hours is estimated based on a group with especially severe symptoms. Such a calculation will overstate the impact of long COVID on the size of the labor force.[13]

One study of how long COVID has affected labor supply that generated headlines when it was released is by Katie Bach (2022b). She concludes that, as of June 2022, long COVID may have reduced labor supply by the equivalent of 2 to 4 million full-time workers, even ignoring any effects on people older than 65.[14] To arrive at her estimates, Bach first cites

13. Sheiner and Salwati (2022) also make this point.

14. This study updates an earlier estimate reported in Bach (2022a) that, as of December 2021, long COVID reduced labor supply by the equivalent of 1.6 million full-time-equivalent workers.

an estimate from the Household Pulse Survey that, as of June 2022, about 16 million Americans age 16–65 were experiencing long COVID symptoms, a number that seems roughly in line with other available estimates. She assumes that, absent long COVID, the labor force participation rate for these people would have been the same as the average for the 16–65 age group (about 75 percent). This also seems reasonable. Bach's final step is to apply alternative estimates of the share of workers with long COVID who leave the labor force and the share who reduce their hours, together with a rough estimate of how large the reduction in hours might have been. This is where the difficulties arise.

Two of the three studies of how long COVID affects labor supply that Bach (2022b) cites seem especially questionable as sources of estimates to be applied to the full long COVID population. Her largest estimate of the long COVID effect is that the equivalent of 4 million people age 16 to 65 were not working because of long COVID. This estimate draws on employment and hours impacts reported by respondents to an online non-probability survey for which people with suspected and confirmed COVID-19 were recruited through COVID-19 support groups and social media (Davis and others 2021). The employment and hours impacts underlying her midrange estimate—that long COVID reduced labor supply by the equivalent of 3 million workers—come from an online non-probability survey conducted by the United Kingdom's Trades Union Congress. The methodology statement for this survey states: "The survey was open between 3 April and 27 May 2021 and was promoted on social media, through affiliated unions and long COVID support groups" (Trades Union Congress 2021). Given the way in which participants were recruited, both surveys seem likely to overrepresent people for whom long COVID was an especially severe problem. Bach's lowest estimate—that COVID-19 reduced labor supply among those age 16–65 by the equivalent of 2 million full-time workers—is based on data collected in the June 2021 wave of the Understanding America Survey (UAS) COVID-19 panel (Ham 2022). Although it is difficult to be certain how representative the respondents to the UAS COVID-19 panel are and the sample in any case included only 193 long COVID sufferers, this is the most plausible of Bach's estimates.[15]

15. While this estimate is more plausible than the others, it is not clear to us how Bach (2022b) translated the UAS estimates into impacts on employment and hours. Ham (2022) reports that, in the UAS data, 25.9 percent of people with long COVID say their work has been "impacted," meaning that they either left work or were working fewer hours, but the survey did not ask separately about effects on the extensive versus the intensive margin of labor supply.

Obtaining accurate estimates of long COVID prevalence and its effects on labor supply admittedly is difficult. Goda and Soltas (2023) produce an estimate using an approach that does not rely on directly identifying long COVID sufferers. They infer long COVID's effects on participation using data on the prevalence of health-related work absences and the relationship of these absences to later labor force withdrawal. This approach assumes, as seems supported by available evidence, that the significant increases in health-related absences from work following the start of the pandemic can be attributed to COVID-19 infections. Another key assumption is that, as also seems to be supported by the data, health-related absences affect labor force withdrawals and reductions in hours similarly whether they are due to COVID-19 or something else.

In their analysis, Goda and Soltas (2023) begin by identifying CPS respondents who missed an entire week of work for health-related reasons. Using the longitudinal structure of the CPS to link these records to other interviews with the same people, Goda and Soltas ask how absences from work affect the probability that a person is working at later points in time. Their baseline estimate is that labor force participation falls by about 7 percentage points following a health-related absence. They also find that, following the health-related absence, those who remain employed reduce their hours and are more likely to shift into part-time jobs.

Goda and Soltas (2023) translate their estimate of how a health-related absence affects subsequent participation into an estimate of the overall effect of COVID-19 on the size of the labor force by combining it with estimates of the number of excess health-related absences during the COVID-19 period. A limitation of the analysis is that people are observed for no more than 14 months after their health-related absence. Goda and Soltas's baseline estimate assumes that the effect on participation in later months continues to decay at the same rate as from months 1 to 14. The more extreme cases that bound this estimate assume either instantaneous complete decay (no effect on participation beginning in month 15) or no further decay (effects on participation that persist at the same level as in month 14). Overall, these calculations imply that, as of June 2022, COVID-19 had reduced the size of the labor force by between 340,000 and 590,000 people, with a baseline estimate of 500,000 people.

As Goda and Soltas (2023) explain, there are at least two reasons why these estimates might be too low. First, the estimates include only people who experienced COVID-19 while employed. Assuming that the likelihood of experiencing a serious bout of COVID-19 is the same for nonworkers as for workers and that COVID-19 reduces the likelihood of a

nonworker entering the labor force by the same percentage as it increases the likelihood of a worker leaving the labor force, accounting for this adds about 50,000 people to Goda and Soltas's baseline participation estimate. Second, some workers could have experienced COVID-19-related absences outside the CPS reference week. Using information on the rate of month-to-month persistence in being absent for health-related reasons to estimate the rate of escape, they estimate that the typical absence lasts a little over three weeks. This implies an adjustment to their baseline participation estimates of about 110,000 people. Together, these adjustments raise Goda and Soltas's baseline estimate of the participation effect by 32 percent to 660,000 people out of the labor force in June 2022 due to earlier having contracted COVID-19, with an implied range of about 450,000 to 780,000 people.[16]

We have built on the analysis by Goda and Soltas (2023) in several ways. First, to assess whether the estimated effect of long COVID has been growing, we extended the time period for the analysis through December 2022. We find somewhat larger long COVID effects as of December 2022 than as of June 2022, though given the nature of these estimates we would be cautious about placing too much weight on the differences. Our baseline estimate for December 2022 is that 620,000 people were out of work as a result of previous excess health-related work absences, with a range from 370,000 to 730,000 people. This is up from a baseline estimate of 560,000 people and a range of 350,000 to 610,000 people for June 2022.[17] Inflating the December numbers to account for absences from work that did not occur during the CPS reference week and for people who were not working at the time of their health-related episode, our baseline estimate for December 2022 is 820,000 people, with a range of 480,000 to 960,000 people.

Second, after replicating Goda and Soltas's (2023) event study estimates of the effects of absences from work on hours in subsequent months, we convert those hours effects into an estimate of worker equivalents lost

16. Goda and Soltas (2023) report estimates of how making the two adjustments just described affects their baseline estimates; we have used those numbers to approximate the corresponding figures for their low and high estimates.

17. Our June 2022 estimates are slightly higher than the original estimates in Goda and Soltas (2023). One reason is that rather than attempting to estimate the decay rate from the data, which proved to be problematic for subgroup analysis, we assume for our baseline estimates that the effects of health-related absences on participation and hours fade out over the three years following the last observation at 14 months after the absence. Another difference in our calculations is that we used weights adjusted for the effect of the CPS population controls introduced in January 2022 and January 2023.

by dividing the implied total loss in hours by average weekly hours as of December 2019. Our estimate assumes that the incidence of weeklong COVID-19 spells is the same for nonworking as for working individuals and that having experienced a weeklong COVID-19 spell affects the later hours of nonworking individuals who become employed in the same way as it affects the later hours of individuals who were employed at the time of their illness. By the same reasoning that Goda and Soltas apply to their labor force effect estimates, we also adjust the figures to account for absences that occurred outside the CPS reference week. Under the same alternative assumptions as used to estimate the participation effects about the decay rate for long COVID effects following the end of the period we are able to observe, our baseline estimate is an hours effect of about 480,000 worker equivalents, with range of 260,000 to 580,000 people.[18] An important caveat, however, is that there is a pre-trend in the event study estimates for hours that the controls included in the model do not eliminate, something that Goda and Soltas also note. The true effect of long COVID on hours thus likely is lower than suggested by these estimates.

Our final extension to the analysis in Goda and Soltas (2023) is to build on their results to investigate how the decline in labor supply attributable to long COVID was distributed across people with different demographic characteristics. We do not find any notable differences by sex or by race and ethnicity. We were especially interested in whether there were differences between people under age 65 and people age 65 and older, as we thought differential long COVID effects might explain the differing pattern of labor force changes between older and younger adults. More specifically, our objective was to determine whether long COVID might help to explain why participation fell more for older adults. Results reported by Goda and Soltas (2023) show (and we confirm) that the effects of a health-related work absence lasting a week or more on subsequent labor force participation are larger for adults age 65 and older than for younger adults. The excess rate of weeklong absences for health reasons during the pandemic period also was larger for older workers. Despite this, reflecting the lower pre-COVID-19 participation rate among those over age 65, the *percentage point* participation rate effect was *smaller* for this older group than for adults age 16–64.

Taking a different approach, Sheiner and Salwati (2022) develop estimates that make use of CPS information on disability status. As was true

18. To translate hours into worker equivalents we assume they work the December 2019 average weekly hours.

of the prevalence of health-related absences from work, the share of people reporting a disability rose noticeably following the start of the pandemic, both absolutely and relative to trend. Using the different labor force participation rates they observe for people with and without a disability both before and after the COVID-19 period, Sheiner and Salwati tease out estimates of the impact of long COVID on overall labor force participation among those age 16–64.

To produce these estimates, Sheiner and Salwati (2022) estimate the number of people disabled due to long COVID as the difference between the number reporting a disability and the number that would have been expected based on the 2017–2019 trend in the disability rate. They assume that, had they not gotten sick, those with long COVID would have had the same participation rate as nondisabled adults of the same age and sex. An assumption also is needed about the evolution of the participation rate for people with an existing non-COVID-19 disability during the COVID-19 period. The simplest assumption is that it would have been the same as in 2019, but Sheiner and Salwati also consider an alternative with a continuation of the rising trend in participation for those with a disability estimated over the 2017–2019 period. Based on these assumptions, they estimate that long COVID reduced the size of the labor force by an average of between 281,000 and 562,000 people over the January–September 2022 period, where the larger number assumes that the participation rates of previously disabled adults continued on their rising trend. An additional complication is that the growing availability of remote work since the pandemic could have allowed more of the existing population with disabilities to enter the labor force. If this has occurred and increased the participation of disabled adults age 45–64 by 5 percent, the total estimated labor force shortfall among those age 16–64 attributable to COVID-19 is larger, in the range of roughly 400,000 to 683,000 people. Sheiner and Salwati also produce an estimate of the effect of long COVID on hours worked by those who remain employed of roughly 20,000 to 39,000 full-time equivalents. This contrasts with the considerably larger effect on hours we obtain based on the approach used by Goda and Soltas (2023), though as noted there is reason to suspect that the latter estimate is biased upward.

Sheiner and Salwati (2022) chose to focus their analysis on adults age 16–64, but the labor market decisions of older adults also may have been affected by long COVID, and we have replicated their calculations including adults age 65 and older. Our calculations make use of 2022 data for the full calendar year rather than for January–September as in Sheiner and Salwati (2022). Adding participation shortfalls among those age 65 and

older raises the estimated long COVID effects; we estimate a range of 324,000 to 858,000 people out of work due to long COVID, where the former assumes the same participation rates for the existing disabled as in 2019 and the latter assumes a continuation of the prior upward trend in participation for that group together with an additional increase in their participation due to growth in the opportunity for remote work.[19]

We also have replicated Sheiner and Salwati's (2022) hours analysis. For consistency with calculations reported elsewhere in the paper, we translate the total implied hours lost into worker equivalents by dividing by average weekly hours as of December 2019. Our original intention had been to include people age 65 and older in these calculations, but we could not obtain sensible results for that group and abandoned the effort. Our estimates are nonetheless somewhat larger than Sheiner and Salwati's estimates, both because our hours denominator is slightly smaller than the 40 hours per week they assume and because, using the same denominator, the hours shortfall in 2022 compared to the pre-pandemic period looks larger when considering data for the full calendar year as opposed to just January–September. Hours are more seasonal than participation and are relatively high during the summer months, which means that comparing estimates for January–September 2022 to estimates for the full 2019 calendar year leads to an understatement in how much hours changed. Even so, the worker equivalent effect of reductions in hours due to long COVID that we estimate remains small compared to the effect of long COVID on participation, just 42,000 to 62,000 workers.

Based on all the available evidence as just described, our best guess is that long COVID may have reduced the size of the labor force as of the end of 2022 by perhaps 710,000 people, a participation rate decline of a bit less than 0.3 percentage point. As already discussed, the presence of a pre-trend in the event study coefficients for the hours estimates leads us to suspect that estimates of the hours effect based on the approach taken by Goda and Soltas (2023) are too large. On the other hand, we are concerned that the estimates of the hours effect based on the approach taken by Sheiner

19. In addition to the modifications mentioned in the text, we also use CPS weights adjusted for the introduction of new population controls in January 2022 and January 2023 rather than the unadjusted weight used in the original calculations in Sheiner and Salwati (2022). The calculations for our high-end estimate implied that, for the 65 and older age group, the number of individuals who left the labor force due to COVID-19 was larger than the number of people in that group who would have been working had they not contracted COVID-19. We use the latter as our high-end estimate of the estimated COVID-19 effect.

and Salwati (2022) are sensitive to noise in the data. We split the difference between the two and place the effect of long COVID on hours at the equivalent of about 270,000 workers.[20]

II.E. Fear of COVID-19

Another explanation offered for the persistent post-pandemic shortfalls in labor supply is that fear of contracting COVID-19 or spreading COVID-19 to a family member kept potential workers on the sidelines. To some extent, this may overlap with the effects of long COVID, in that those suffering from long COVID may have been more fearful of contracting COVID-19 again. We know of two ongoing surveys that have asked questions relevant to understanding how fear of COVID-19 has affected participation—the Survey of Working Arrangements and Attitudes (SWAA) and the Census Bureau's Household Pulse Survey (HPS). The SWAA is an online non-probability survey administered monthly since May 2020 to Americans age 20–64 (Barrero, Bloom, and Davis 2022, 2023). The SWAA sample initially was restricted to individuals with significant prior year work attachment; this restriction was dropped beginning in June 2022.[21] The survey has included direct questions about fear of COVID-19 and other infectious diseases as a reason for nonparticipation since February 2022.[22] The HPS is an experimental probability-based survey first fielded by the Census Bureau at the end of April 2020. The sample for the HPS is drawn from households on the Census Bureau's Master Address File (MAF) that could be matched to a phone number (available for 88 percent of addresses) and/or email address (available for 80 percent of addresses). Data were collected for adults age 18 and older using an online platform (Fields and others 2020). A question on reasons for nonparticipation was included in each of the forty-seven waves of the HPS from June 2020 through December 2022.

20. The 710,000 and 270,000 numbers are the averages of our baseline estimates based on the approach used by Goda and Soltas (2023) and the midpoints of the ranges we obtain using the approach used by Sheiner and Salwati (2022).

21. More specifically, from May 2020 to March 2021, the requirement for inclusion in the SWAA sample was earnings of at least $20,000 in 2019. From April to September 2021, the earnings threshold transitioned to $10,000 in 2019, and from January to May 2022, it transitioned to $10,000 during the prior year, before being dropped in June 2022.

22. The SWAA includes a separate question about plans to continue social distancing, also analyzed by Barrero, Bloom, and Davis (2022, 2023). They find that people who plan to continue social distancing also are more likely not to be working, but that finding is more difficult to interpret than the findings based on answers to questions about fear of COVID-19 as a factor in the decision to participate.

Jose Maria Barrero, Nicholas Bloom, and Steven J. Davis are the research team that has looked most directly at the effect of COVID-19 fears on labor force participation. In an October 2022 working paper (Barrero, Bloom, and Davis 2022), they reported estimates based on responses from SWAA sample members with significant prior year earnings who were not currently working or looking for work to the question: "Are worries about catching COVID or other infectious diseases a factor in your decision not to seek work at this time?" The three response options were: "Yes, the main reason"; "Yes, a secondary reason"; and "No." Their estimates assume that all of those who said their main reason for not working or seeking work was fear of COVID-19 and that half of those who said it was a secondary reason otherwise would have been in the labor force. Under these assumptions, they estimate that COVID-19 fears reduced labor force participation in their sample from February through July 2022 by 2.0 percentage points.[23]

One limitation of the initial estimates in Barrero, Bloom, and Davis (2022) is that the SWAA sample on which they are based is restricted to people with significant prior year work attachment. Beginning in June 2022, the SWAA sample was expanded to include all adults age 20–64. Barrero, Bloom, and Davis kindly shared with us estimates for the June–December 2022 period based on the survey question just described, both for the set of people satisfying the original SWAA sample selection criteria and for the expanded sample. Assuming that half of those citing COVID-19 fears as a secondary reason for nonparticipation otherwise would have been in the labor force, the estimated impact of COVID-19 fears for December 2022 in the restricted sample was 1.2 percentage points, down from 2.0 percentage points earlier in the year. For the more inclusive sample, it was 1.9 percentage points in December 2022.

To put this number in context, the overall labor force participation rate for adults age 20–64 was just 0.2 percentage point lower in December 2022 than it had been in December 2019 (77.4 percent compared to 77.6 percent). The 1.9 percentage point estimate thus implies that, absent the fear of COVID-19, participation among adults age 20–64 would have been 1.7 percentage points *higher* in December 2022 than it had been in December 2019. Given the similarity of overall labor market conditions in those

23. Although the SWAA question is worded more generally to encompass "worries about catching COVID or other infectious diseases," during this period it seems reasonable to think that the infectious disease most people have in mind is COVID-19. For consistency with the data for other months, Barrero, Bloom, and Davis (2022) dropped June 2022 and July 2022 respondents who did not satisfy the prior earnings requirement in place from February 2022 through May 2022 from the estimation sample.

months and that participation generally has been trending downward, we do not find this plausible.

A number of questions can be raised about the estimates reported by Barrero, Bloom, and Davis (2022), most of which the authors themselves have investigated and, where possible, addressed. One obvious question is whether participants in the pre-recruited online panels used for the SWAA are representative of the target population. Barrero, Bloom, and Davis report that, in a sample that imposed a prior earnings requirement and using answers to a different question about the main reason a person is not working, the SWAA estimate of the share of people saying they were not working because of COVID-19 fears was about 0.6 percentage point higher than when the same question was asked in the HPS.[24] Because of the relatively small SWAA sample size, this difference is not statistically significant, but it may suggest issues with the representativeness of the SWAA sample.

In addition, the way the original SWAA question was structured seems likely to have inflated the estimated effect of COVID-19 fears in the initial analysis by Barrero, Bloom, and Davis (2022). The original SWAA question asks directly whether COVID-19 fears are a factor in not working or looking for work. This construction of the question is potentially subject to acquiescence bias—the tendency of survey respondents to agree with a survey statement whether or not it reflects their true opinion. The share of people who said fear of COVID-19 was the main reason they were out of the labor force was considerably larger than the share who chose fear of COVID-19 as their main reason for not working when responding to the HPS question that offers respondents a menu of possible reasons for nonparticipation (Barrero, Bloom, and Davis 2022).[25]

Beginning in October 2022, the SWAA has included new questions that address this concern. These questions ask in turn for the main reason and the second most important reason a person is not working or looking for work. Respondents are given a menu of possible responses to each of these questions, one of which is: "I worry about catching COVID or other infectious diseases." As reported by Barrero, Bloom, and Davis (2023), an update to Barrero, Bloom, and Davis (2022), when the answers to these questions are translated into participation effects, assuming a 50 percent impact for people saying COVID-19 fears were a secondary factor, the estimates are considerably smaller. For the sample with no prior earnings

24. The difference between the two estimates rises to 0.8 percentage point when the SWAA sample is restricted to respondents who gave an acceptable answer to included "attention check" questions.

restrictions, the new question yields an estimated effect of COVID-19 fears on participation for December 2022 of 1.0 percentage point, compared to 1.9 percentage points based on the original question.

A final comment about the estimates by Barrero, Bloom, and Davis (2022, 2023) concerns the translation of the responses to the fear of COVID-19 question into participation impacts. The assumption underlying all of the featured estimates is that half of those who cite COVID-19 fears as a secondary reason for nonparticipation otherwise would have been working or looking for work. One might doubt whether the marginal effect on participation of concerns about COVID-19 among people who have already said their main reason for not working is something else is this large. If the true effect for this group is smaller, the overall impact on participation would, of course, also be smaller. As an illustration, the numbers reported in table 6 of Barrero, Bloom, and Davis (2023) imply that, if it were assumed that 10 percent rather than 50 percent of those who give COVID-19 fears as a secondary reason for nonparticipation otherwise would have been in the labor force, the estimated effect of COVID-19 fears on the December 2022 participation rate would have been 0.4 percentage point rather than 1.0 percentage point.

The HPS is a second source of information on how COVID-19 fears have affected participation. Although the HPS also may suffer from representation problems—its response rate has averaged under 10 percent—it is a probability-based sample and low survey response rates do not necessarily imply bias in survey estimates (Groves and Peytcheva 2008). Some research has found that probability samples tend to produce more accurate estimates than nonprobability samples even when the response rates to the former are low (Yeager and others 2011), though the low HPS response rate nonetheless suggests caution in interpreting the estimates.

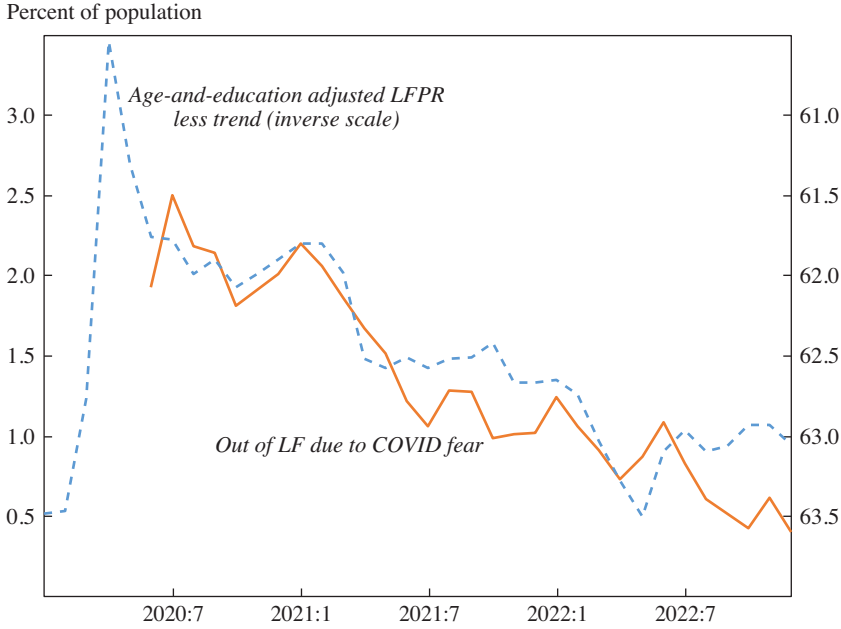
The HPS question about nonparticipation asks adults age 18 and older who were not working during the survey reference week the main reason they were not employed.[26] "I was concerned about getting or spreading the coronavirus" is one of roughly a dozen response options.[27] Figure 5 shows

25. As mentioned above, the SWAA question is also worded more generally to encompass "worries about catching COVID or other infectious diseases," but the mention of other infectious diseases seems unlikely to account for the much higher number of "yes" responses.

26. The question is asked of everyone who is not working rather than only of those who are out of the labor force, but if someone gives fear of COVID-19 as the reason they are not employed, it is likely this fear would also prevent them from actively seeking work.

27. The exact list of response options has varied slightly over time, but the option related to fear of getting or spreading the coronavirus has been included on the list in every wave in which the question was asked.

Figure 5. Labor Force Participation Rate Adjusted for Age, Education, and Trend and Percentage of Population out of Labor Force Due to COVID-19 Fears, January 2020 to December 2022



Sources: Authors' tabulations of Current Population Survey and Household Pulse Survey data.

Note: Labor force participation series constructed using monthly CPS data downloaded from the IPUMS data set (Flood and others 2022) and reweighted to wedge effects of new population controls introduced in January 2022 and January 2023 backward. Monthly participation estimates seasonally adjusted using X-12 seasonal adjustment command in Eviews, applying multiplicative X-11 method and auto X-12 seasonal and trend filter. Participation rate is for persons age 16 and older; share not working due to fear of COVID is for persons age 18 and older.

the HPS estimate of the share of the population age 18 and older who said that fear of COVID-19 was their main reason for not working and (on an inverse scale) the overall labor force participation rate adjusted for changes in demographic mix and pre-pandemic within-group trends. The HPS estimates are for the survey wave closest in time to each month's CPS reference week or, in a few cases, are interpolations based on the values for adjacent months. According to the HPS estimates, the share of people 18 and older who said they were not working because of COVID-19 fears peaked at 2.5 percent in July 2020. By February 2022, it had fallen to 1.1 percent, and by December 2022 to 0.4 percent. The changes in this series and in the labor force participation rate are of similar magnitude, and they have moved remarkably closely together over time (the Pearson

correlation between them is –0.91). This is consistent both with COVID-19 fears having been a more significant part of the explanation for the shortfall in participation earlier in the pandemic and also with their having become much less important over time.

While the overall relationship between the percentage of people saying they are not working because of COVID-19 fears and the adjusted participation rate is very strong, it breaks down when the data are disaggregated by demographic group. As discussed in the first section of the paper, the age group with the largest unanticipated declines in participation since the start of the pandemic has been adults age 65 and older, but this is also the age group least likely to say they are out of the labor force due to COVID-19 fears. We initially found this somewhat counterintuitive, since COVID-19 risks are well known to be more serious for older adults. The responses regarding fear of COVID-19 may be less meaningful for older adults because they are more likely to choose retirement as the main reason they are out of the labor force even if COVID-19 fears contributed to their choosing not to work. In the data by race and ethnicity, as documented earlier, after adjusting for age, education, and trend effects, overall participation has been below expectations for non-Hispanic white adults but not for non-Hispanic Black adults and only very slightly so for Hispanic adults. The share of Hispanic adults and non-Hispanic Black adults citing fear of COVID-19 as a reason for being out of the labor force, however, has been significantly higher than the share of white non-Hispanic adults. The fear of COVID-19 information in the HPS is, of course, collected only from those who are not working. To the extent that COVID-19 concerns are in fact generally greater among groups where a larger percentage of people are out of the labor force because of them, however, it may be that the necessity of working—and the opportunity to do so in a very tight labor market—has pushed people in some groups that on average have more limited resources back into the labor market despite real health concerns.

There are, of course, reasons to be cautious in interpreting these numbers. Given the survey's very low response rate, one might wonder about the representativeness of the HPS sample. Because the HPS question asks only about respondents' main reason for not working, it may miss some people for whom fear of COVID-19 is a contributing factor, though it is somewhat unclear how important this might be. On the other hand, the responses to the question about why people are not working may be affected by social desirability bias. Some people may feel it puts them in a better light to say the reason they are not working is fear of COVID-19

rather than, for example, that they did not want to be working at this time, leading the estimated effect of COVID-19 fears on labor force participation to be exaggerated. In addition, fear of COVID-19 also may overlap to some extent with having experienced long COVID. Although this is necessarily a guess—albeit a guess informed by our assessment of available data—if forced to assign a number, we would peg the effect of COVID-19 fears on labor force participation as of December 2022 at perhaps 0.3 percentage point or about 800,000 people.

III. Conclusion

The labor market changed significantly over the three years following the start of the pandemic. As of December 2022, the labor force participation rate was about 1.1 percentage points below its December 2019 level. This is a shortfall of about 3 million workers. The 0.6 hour reduction in average weekly hours over the same period contributed an additional labor supply shortfall that is the equivalent of about another 2.6 million workers. Our goal in this paper has been to better understand what explains these changes.

Table 3 summarizes our assessment of the changes in participation and average weekly hours between December 2019 and December 2022, together with the factors we believe contributed to those changes. Panel A of the table reports estimates for the decline in the labor force participation rate; panel B translates those declines into thousands of people; and panel C presents estimates for the decline in average weekly hours in terms of their equivalent in thousands of people. As we hope is clear from the discussion earlier in the paper of the evidence on which these estimates are based, the numbers in the table are very much in the nature of guesstimates. Although we do not attach great confidence to the exact magnitudes reported, we are more confident in our general conclusions regarding the orders of magnitude of the various effects.

As shown in the table, much of the decline in labor force participation over the three years ending in December 2022 should have been anticipated even absent the pandemic. Exactly how much change should have been anticipated depends on what one believes to be the relevant counterfactual. If the labor force participation rate would have evolved after December 2019 based solely on demographic factors—specifically, changes in the age and education composition of the population—participation at the end of December 2022 was a little less than 0.6 percentage point or about

Table 3. Explaining the Post-pandemic Declines in Labor Force Participation and Worker-Equivalent Average Hours, December 2019 to December 2022

	Counterfactual: Demographic adjustment only	*Counterfactual: Demographic and trend adjustment*
Panel A: Total labor force participation rate decline through December 2022	1.13	1.13
Anticipated decline based on chosen counterfactual	0.55	0.84
Unanticipated decline based on chosen counterfactual	0.58	0.29
Selected pandemic-related factors		
Healthier household balance sheets	0.00	0.00
Active COVID-19 infections	0.00	0.00
Long COVID	0.27	0.27
Fear of COVID-19	0.30	0.30
Residual unexplained decline	0.01	–0.28
Panel B: Total labor force decline through December 2022 (thousands)	3,000	3,000
Anticipated decline based on chosen counterfactual	1,460	2,230
Unanticipated decline based on chosen counterfactual	1,540	770
Selected pandemic-related factors		
Healthier household balance sheets	0	0
Active COVID-19 infections	0	0
Long COVID	710	710
Fear of COVID-19	800	800
Residual unexplained decline	30	–740
Panel C: Total worker-equivalent hours decline through December 2022 (thousands)	2,610	2,610
Anticipated decline based on chosen counterfactual	120	300
Unanticipated decline based on chosen counterfactual	2,490	2,310
Selected pandemic-related factors		
Healthier household balance sheets	0	0
Active COVID-19 infections	710	710
Long COVID	270	270
Fear of COVID-19	0	0
Residual unexplained decline	1,510	1,330

Source: Authors' estimates.
Note: Basis for rough effect size estimates described in text. Person counts rounded to nearest ten thousand.

1.5 million people below where we would have expected it to be. If the evolution of the participation rate also would have reflected the continuation of preexisting within-group trends, the unexplained shortfall is somewhat less than 0.3 percentage point or about 0.8 million people.

In addition to these anticipatable factors, we believe that both long COVID and the fear of COVID-19 put downward pressure on the post-pandemic participation rate, though in both cases we estimate the magnitude

of the effect to be considerably smaller than some previous analyses have suggested. The demographic factors incorporated in our first counterfactual together with long COVID and fear of COVID-19 fully explain the decline in the participation rate from December 2019 through December 2022; those factors plus the preexisting within-group trends incorporated in the second counterfactual more than fully explain it. Put somewhat differently, under the second counterfactual, had the labor market been as strong as it was in December 2022 and had the pandemic not occurred, the labor force participation rate would have been almost 0.3 percentage point higher than it was in December 2019. Although the unemployment rate was comparable in December 2022 to what it had been in December 2019, as discussed earlier in the paper, other indicators including the job vacancy rate provide reason to suspect the December 2022 labor market was even tighter. This makes it plausible that, absent the pandemic and with the tightness of the labor market as it was in December 2022, demographic- and trend-adjusted participation in December 2022 could have been a bit above its December 2019 level.

We have been less successful in explaining why the changes in labor force participation have differed across demographic groups. The largest reductions in participation relative to expectations occurred among older adults and especially the most-educated older adults. The unanticipated shortfall in participation was greater among white non-Hispanic adults than among Hispanic adults, and participation among Black non-Hispanic adults actually rose slightly. We have not found direct evidence that allows us to account for the differences we observe. Many older adults would have been exiting the labor force shortly regardless, and those with means likely have had more flexibility about exactly when that transition occurred. Even if they did not contract COVID-19 and were not driven by fear of COVID-19, living through the pandemic may have led some to reevaluate how long they wanted to continue working. The finding reported by Favilukis and Li (2023) that participation declined more among older homeowners who benefited from rising house prices is consistent with the speculation that some older adults rethought their priorities. White non-Hispanic adults as a group have greater accumulated wealth than Black non-Hispanic or Hispanic adults and the decision not to work could have been more financially feasible for them.

Panel C of table 3 tells a rather different story about average weekly hours. Neither demographic changes nor preexisting trends can account for much of the drop in average weekly hours we observe between December 2019 and December 2022. Time missed due to acute COVID-19 infections and

reductions in hours resulting from long COVID both have been a factor, but in our estimation these account for less than 40 percent of the hours decline. In contrast to our conclusion with regard to participation, we do not believe that fear of COVID-19 helps to explain the decline in hours. More of the decline in hours thus remains to be explained.

Average weekly hours as measured in the CPS are quite cyclical, and it is possible that, given time, they will recover to pre-pandemic levels. It also is possible, however, that the lower level of hours reflects a more permanent reevaluation regarding the balance people wish to strike between their work and personal lives. The media are full of stories about "quiet quitting" (Telford 2022; Rosalsky and Selyukh 2022) and professionals who are opting to step back from demanding working schedules (Krueger 2022). Consistent with this narrative, Faberman, Mueller, and Şahin (2022) provide evidence that desired hours of work fell during the pandemic. To the extent this accounts for the reduction in hours we observe, it could be much longer lasting.

ACKNOWLEDGMENTS We thank Stephanie Aaronson, Steven Davis, Jan Eberly, Yongseok Shin, James Stock, participants in a seminar at the Harris School of Public Policy, and participants in the *BPEA* meeting for helpful comments on earlier drafts of this paper. We are grateful to Ayşegül Şahin, Bart Hobijn, Gopi Shah Goda, Evan Soltas, Louise Sheiner, Nasiha Salwati, Jose Maria Barrero, Nicholas Bloom, and Steven Davis for sharing data and code with us.

References

Aaronson, Daniel, Jonathan Davis, and Luojia Hu. 2012. "Explaining the Decline in the U.S. Labor Force Participation Rate." Chicago Fed Letter 296.

Aaronson, Stephanie, Tomaz Cajner, Bruce Fallick, Felix Galbis-Reig, Christopher Smith, and William L. 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 F. Pingle, and William L. 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: 69–154.

Abraham, Katharine G., Brad Hershbein, Susan N. Houseman, and Beth Truesdale. 2023. "The Independent Contractor Workforce: New Evidence on Its Size and Composition and Ways to Improve Its Measurement in Household Surveys." Working Paper 30997. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30997.

Abraham, Katharine G., and Melissa S. Kearney. 2020. "Explaining the Decline in the US Employment-to-Population Ratio: A Review of the Evidence." *Journal of Economic Literature* 58, no. 3: 585–643.

Aiyegbusi, Olalekan Lee, Sarah E. Hughes, Grace Turner, Samantha Cruz Rivera, Christel McMullan, Joht Singh Chandan, and others. 2021. "Symptoms, Complications and Management of Long COVID: A Review." *Journal of the Royal Society of Medicine* 114, no. 9: 428–42.

Bach, Katie. 2022a. *Is 'Long Covid' Worsening the Labor Shortage?* Washington: Brookings Institution. https://www.brookings.edu/research/is-long-covid-worsening-the-labor-shortage/.

Bach, Katie. 2022b. *New Data Shows Long Covid Is Keeping as Many as 4 Million People Out of Work*. Washington: Brookings Institution. https://www.brookings.edu/research/new-data-shows-long-covid-is-keeping-as-many-as-4-million-people-out-of-work/.

Barnes, Mitchell, Wendy Edelberg, Sara Estep, and Moriah Macklin. 2022. *Bolstered Balance Sheets: Assessing Household Finances since 2019*. Washington: Hamilton Project, Brookings Institution. https://www.brookings.edu/research/bolstered-balance-sheets-assessing-household-finances-since-2019/.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2022. "Long Social Distancing." Working Paper 30568. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30568.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2023. "Long Social Distancing." Working Paper. Chicago: Becker Friedman Institute, University of Chicago. https://bfi.uchicago.edu/working-paper/long-social-distancing/.

Bauer, Lauren, Wendy Edelberg, Sara Estep, and Brad Hershbein. 2023. "Who's Missing from the Post-Pandemic Labor Force?" Washington: Brookings Institution. https://www.brookings.edu/2023/04/04/whos-missing-from-the-post-pandemic-labor-force/.

Bosworth, Barry P., and Gary Burtless. 2010. *Recessions, Wealth Destruction, and the Timing of Retirement*. Washington: Brookings Institution. https://www.brookings.edu/research/recessions-wealth-destruction-and-the-timing-of-retirement/.

Cesarini, David, Erik Lindqvist, Matthew J. Notowidigdo, and Robert Ostling. 2017. "The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish Lotteries." *American Economic Review* 107, no. 12: 3917–46.

Coile, Courtney C. 2022. "Retirement during the COVID-19 Pandemic." Paper presented at the Federal Reserve Bank of Boston 66th Economic Conference, November 18–19. https://www.bostonfed.org/-/media/Documents/events/2022/labor-markets/papers/retirement-during-the-covid19-pandemic-coile.pdf.

Coile, Courtney C., and Phillip B. Levine. 2011. "The Market Crash and Mass Layoffs: How the Current Economic Crisis May Affect Retirement." *B.E. Journal of Economic Analysis and Policy* 11, no. 1: 1–42.

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

Davis, Hannah E., Gina S. Assaf, Lisa McCorkell, Hannah Wei, Ryan J. Low, Yochai Re'em, Signe Redfield, Jared P. Austin, and Athena Akrami. 2021. "Characterizing Long COVID in an International Cohort: 7 Months of Symptoms and Their Impact." *eClinicalMedicine* 38:101019.

Domash, Alex, and Lawrence H. Summers. 2022. "How Tight Are U.S. Labor Markets?" Working Paper 29739. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29739.

Faberman, R. Jason, Andreas I. Mueller, and Ayşegül Şahin. 2022. "Has the Willingness to Work Fallen during the Covid Pandemic?" Working Paper 29784. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29784.

Farnham, Martin, and Purvi Sevak. 2016. "Housing Wealth and Retirement Timing." *CESifo Economic Studies* 62, no. 1: 26–46.

Favilukis, Jack, and Gen Li. 2023. "The Great Resignation Was Caused by the COVID-19 Housing Boom." Social Science Research Network, January 25. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4335860.

Ferguson, Stephanie. 2023. "Understanding America's Labor Shortage." US Chamber of Commerce, July 13. https://www.uschamber.com/workforce/understanding-americas-labor-shortage.

Fields, J. F., J. Hunter-Childs, A. Tersine, J. Sisson, E. Parker, V. Velkoff, C. Logan, and H. Shin. 2020. "2020 Household Pulse Survey: Interagency Federal Statistical Rapid Response Survey to Measure Effects of the Coronavirus (COVID-19) Pandemic on the United States Household Population." Working Paper. Washington: US Census Bureau. https://www2.census.gov/programs-surveys/demo/technical-documentation/hhp/2020_HPS_Background.pdf.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, and Michael Westberry. 2022. *Integrated Public Use Microdata Series, Current*

Population Survey: Version 10.0 [dataset]. Minneapolis: University of Minnesota. https://doi.org/10.18128/D030.V10.0.

Goda, Gopi Shah, John B. Shoven, and Sita Nataraj Slavov. 2011. "What Explains Changes in Retirement Plans during the Great Recession?" *American Economic Review* 101, no. 3: 29–34.

Goda, Gopi Shah, and Evan J. Soltas. 2023. "The Impacts of Covid-19 Absences on Workers." *Journal of Public Economics* 222:104889.

Groves, Robert M., and Emilia Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72, no. 2: 167–89.

Ham, Dasom I. 2022. "Long-Haulers and Labor Market Outcomes." Working Paper 60. Federal Reserve Bank of Minneapolis, Opportunity and Inclusive Growth Institute. https://www.minneapolisfed.org/research/institute-working-papers/long-haulers-and-labor-market-outcomes.

Hobijn, Bart, and Ayşegül Şahin. 2022. "'Missing' Workers and 'Missing' Jobs since the Pandemic." Paper prepared for the Federal Reserve Bank of Boston Conference on Labor Markets during and after the Pandemic. https://www.bostonfed.org/-/media/Documents/events/2022/labor-markets/papers/missing-workers-and-missing-jobs-since-the-pandemic-hobijn-sahin.pdf.

Hornstein, Andreas, and Marianna Kudlyak. 2019. "Aggregate Labor Force Participation and Unemployment and Demographic Trends." Working Paper 19-08. Federal Reserve Bank of Richmond. https://www.richmondfed.org/publications/research/working_papers/2019/wp_19-08.

Krueger, Alyson. 2022. "Endless Summer Fridays." *New York Times*, April 1. https://www.nytimes.com/2022/04/01/style/four-day-workweek-friday.html.

Lee, Dain, Jinhyeok Park, and Yongseok Shin. 2023. "Where Are the Workers? From Great Resignation to Quiet Quitting." Working Paper 30833. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30833.

Mitchell, Josh, Lauren Weber, and Sarah Chaney Cambon. 2021. "4.3 Million Workers Are Missing. Where Did They Go?" *Wall Street Journal*, October 14. https://www.wsj.com/articles/labor-shortage-missing-workers-jobs-pay-raises-economy-11634224519.

Montes, Joshua. 2018. "CBO's Projection of Labor Force Participation Rates." Working Paper 2018-04. Washington: Congressional Budget Office. https://www.cbo.gov/publication/53616.

Montes, Joshua, Christopher Smith, and Juliana Dajon. 2022. "'The Great Retirement Boom': The Pandemic-Era Surge in Retirements and Implications for Future Labor Force Participation." Finance and Economics Discussion Series 2022-081. Washington: Board of Governors of the Federal Reserve System. https://doi.org/10.17016/FEDS.2022.081.

Morath, Eric, and Te-Ping Chen. 2020. "As $600-a-Week Jobless Aid Nears End, Congress Faces a Quandary." *Wall Street Journal*, July 21. https://www.wsj.com/articles/as-600-a-week-jobless-aid-nears-end-congress-faces-a-quandary-11595151000.

Parker, Jonathan A., Jake Schild, Laura Erhard, and David S. Johnson. 2022. "Economic Impact Payments and Household Spending during the Pandemic." *Brookings Papers on Economic Activity*, Fall, 81–130.

Powell, Jerome H. 2022. "Inflation and the Labor Market." Speech delivered at the Hutchins Center on Fiscal and Monetary Policy, Brookings Institution, November 30. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/newsevents/speech/powell20221130a.htm.

Rosalsky, Greg, and Alina Selyukh. 2022. "The Economics behind 'Quiet Quitting'—and What We Should Call It Instead." *Planet Money*, NPR, September 13. https://www.npr.org/sections/money/2022/09/13/1122059402/the-economics-behind-quiet-quitting-and-what-we-should-call-it-instead.

Sheiner, Louise, and Nasiha Salwati. 2022. "How Much Is Long COVID Reducing Labor Force Participation? Not Much (So Far)." Working Paper 80. Washington: Hutchins Center on Fiscal and Monetary Policy, Brookings Institution. https://www.brookings.edu/research/how-much-is-long-covid-reducing-labor-force-participation-not-much-so-far/.

Telford, Taylor. 2022. "'Quiet Quitting' Isn't Really about Quitting. Here Are the Signs." *Washington Post*, August 21. https://www.washingtonpost.com/business/2022/08/21/quiet-quitting-what-to-know/.

Trades Union Congress. 2021. *Workers' Experiences of Long Covid: A TUC Report.* London. https://www.tuc.org.uk/sites/default/files/2021-06/Formatted%20version%20of%20Long%20Covid%20report%20-%20v1.3.pdf.

US Bureau of Labor Statistics. 2021. "Adjustments to Household Survey Population Estimates in January 2021." https://www.bls.gov/cps/population-control-adjustments-2021.pdf.

US Bureau of Labor Statistics. 2022a. "Adjustments to Household Survey Population Estimates in January 2022." https://www.bls.gov/cps/population-control-adjustments-2022.pdf.

US Bureau of Labor Statistics. 2022b. "Employment Projections: Calculation." In *Handbook of Methods*. https://www.bls.gov/opub/hom/emp/calculation.htm.

US Bureau of Labor Statistics. 2023. "Adjustments to Household Survey Population Estimates in January 2023." https://www.bls.gov/web/empsit/cps-pop-control-adjustments.pdf.

Van Zandweghe, Willem. 2012. "Interpreting the Recent Decline in Labor Force Participation." *Economic Review* 97, no. 1: 5–34.

Yeager, David S., Jon A. Krosnick, LinChiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang. 2011. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-probability Samples." *Public Opinion Quarterly* 75, no. 4: 709–47.

Comments and Discussion

COMMENT BY

STEPHANIE AARONSON I am very happy to have the opportunity to discuss this paper by Katharine Abraham and Lea Rendell. As the authors point out, the post-pandemic labor market has been characterized by an aggregate labor force participation rate (LFPR) that is well below pre-pandemic levels and declining hours of work for those with jobs. Understanding whether these phenomena are driven by temporary factors or whether they are likely to persist is critical both for monetary policymakers trying to estimate the cyclical position of the economy and for fiscal policymakers concerned with, among other issues, projecting tax collections and the future of the Social Security trust funds, which depend importantly on the size of the labor force and how much people are working.

In their paper, Abraham and Rendell attempt to answer this question by focusing primarily on the LFPR, although they also do some analysis of the workweek. They start by decomposing the changes in participation and the workweek into a component that can be explained by demographics and preexisting trends and an unexplained residual. Then they look for possible explanations for the unexplained component, focusing on factors that have arisen since the pandemic, including the large increase in the social safety net and household wealth, which improved household balance sheets, as well as the increased incidence of illness associated with COVID-19 itself. The extent to which the "missing workers" present a problem for policymakers depends on understanding whether current behavior represents a break from the past and whether it is likely to persist. My goal in this discussion, therefore, is to provide an alternative benchmark by which to

Brookings Papers on Economic Activity, Spring 2023: 49–75 © 2024 The Brookings Institution.

judge the current behavior of the participation rate, to provide some additional evidence on the likely persistence of recent trends, and to identify some areas for future research.

For their decomposition, Abraham and Rendell use a shift-share analysis to decompose the changes in the participation rate and workweek from February 2020, just prior to the pandemic, to December 2022. Their explanatory variables include age and education, and in one specification, they also control for pre-pandemic trends within these age and education groups. They calculate the trends using annual data from 2000 to 2019. Since there are indeed long-standing trends within groups—particularly the half-century-long decline in the participation rate among prime-age men—this is my preferred specification, and it is the one that is closest in spirit to my own analysis.

In order to make the comparison, the authors also make an assumption about the cyclical position of the economy. For their analysis, they assume that, given that the unemployment rate is at about the same level as it was pre-pandemic, the cyclical position is also the same. On the one hand, there is reason to think the labor market may have been tighter in winter 2019: by the time of the pandemic, the unemployment rate had been at or below 5 percent for four years and at or below 4 percent for two years. As of the end of 2022, the unemployment rate has been below 5 percent for a bit over two years and below 4 percent for a year. Given the long lags with which the participation rate responds to the economy, participation could have been cyclically stronger prior to the pandemic. On the other hand, measures of labor market tightness other than the unemployment rate, including the high ratio of vacancies to unemployment and the rapid wage growth relative to the pre-pandemic period, suggest the possibility that the labor market is tighter now than it was in early 2020. So, without a clearer benchmark, it is difficult to know whether their assumption that the cyclical position of the economy is the same is a good one.

Altogether Abraham and Rendell find that the participation rate is 1.1 percentage points lower in December 2022 than in December 2019, and most of the decline can be explained. About half the decline is attributable to changing demographics, as the aging of the population imparts a downward trend that is only partially offset by rising educational attainment. On top of that, about 0.3 percentage point is attributable to preexisting trends within groups. This leaves about 0.3 percentage point unexplained.

To check the robustness of this result, I performed my own decomposition, based on the model of the LFPR that I developed with colleagues, then at the Federal Reserve Board, which we first presented at *BPEA* in 2006 and

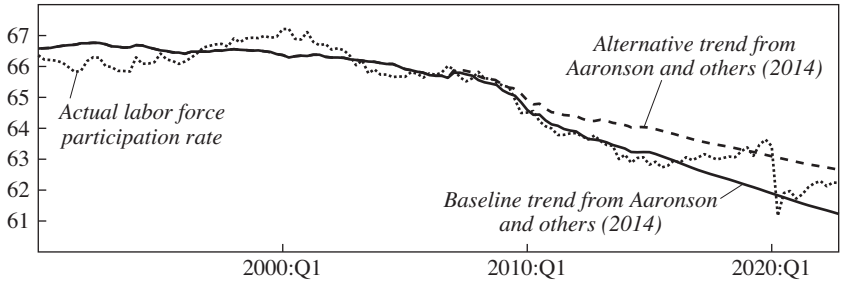
which we reexamined in a subsequent 2014 *BPEA* paper (Aaronson and others 2006, 2014). For this exercise, I did not reestimate the model, but I did crudely update the resulting trends. Similar to Abraham and Rendell, I recalculated the participation rate from the micro data using revised weights that account for the new population controls introduced in January 2022.[1] The data are at a quarterly frequency and were seasonally adjusted using factors from the Bureau of Labor Statistics (BLS). I also adjusted the 2014 model trends to account for the new population controls. I made this adjustment on the basis that if the model had known the population mix present in the new population controls, the level of the trend would have reflected it. To implement the adjustment, I assumed that the 0.3 percentage point increase to the level of the participation rate in January 2022 occurred linearly starting in April 2010. I estimated the cyclical component of the participation rate by regression, using as the measure of labor market slack the unemployment rate gap computed by subtracting the Congressional Budget Office (CBO)'s noncyclical unemployment rate from the aggregate unemployment rate.[2] Research done since we wrote the 2014 paper suggests that the participation rate reacts with very long lags to changes in business cycle conditions—up to four years—and for the analysis I use sixteen lags of the quarterly data.

Figure 1 shows the current participation rate (the dotted line) along with the baseline trend from the model (the solid line) and one of the model's alternative trends (the dashed line). The baseline trend appears to be too low. The level of the participation rate just prior to the pandemic is much higher above trend than can be explained even by the strong labor market that existed at that time. In addition, as I will show later in the discussion, it appears that we were mistaken in our assumption that the downward trend in participation among more recent cohorts would continue.

1. To adjust the micro data using the new population controls I used weights provided by John Coglianese and Christopher Nekarda, which were calculated using the IPUMS CPS and data from the US Census Bureau. The population controls were also updated in February 2023, and, unusually, the Census Bureau will be revising the data back to April 2010. However, at the time of this writing, the data necessary to revise the micro data have not been made publicly available, and since the revision to the level is relatively small (about 0.1 percentage point), I ignore the revision for this analysis. Abraham and Rendell adjust for both revisions to the population controls and the results are similar.

2. For a discussion of using the unemployment rate gap as a measure of slack in participation rate equations, see Aaronson and others (2014, 236–37). The CBO estimate of the noncyclical rate of unemployment does not account for any short-term movements in the natural rate due to dislocation from the pandemic, as, for instance, suggested by the dramatic rise in the ratio of the vacancies to unemployment.

Figure 1. Labor Force Participation Rate and Selected Trends, 1990–2022



Sources: Author's calculations using data from the Current Population Survey, the US Census Bureau, the Bureau of Labor Statistics, and Aaronson and others (2014).

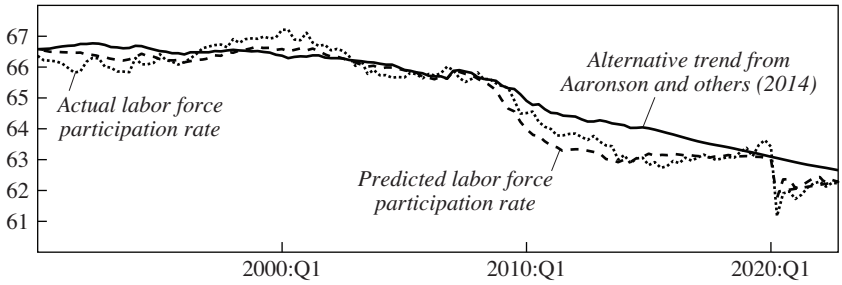
Note: The labor force participation rate is calculated from the Current Population Survey using weights, provided by John Coglianese and Christopher Nekarda, that reflect the US Census Bureau's Vintage 2021 population estimates. The data were seasonally adjusted using factors from the Bureau of Labor Statistics (actual labor force participation rate). The trends have also been adjusted to account for the Vintage 2021 population estimates.

The alternative trend appears to fit the data better. This trend roughly extrapolates forward the level of participation of the last fifteen cohorts, rather than extrapolating forward the downward trend. As I will show, this better matches the flattening out of participation among young people. In addition, this trend lines up well with the cyclical behavior of the participation rate.

Figure 2 shows the participation rate (again the dotted line) against this alternative trend (the solid line) and a predicted value for participation rate based on the trend and its estimated cyclical behavior (the dashed line). As of 2022:Q4, the actual participation rate was in line with the predicted value—in other words, participation was just where you would expect relative to trend, given the cyclical position of the economy. By this measure, there is no unexplained shortfall in the participation rate. That said, given the confidence intervals around the estimates of the trend participation rate, the predicted natural rate, and the unemployment rate gap, this result is not meaningfully different from that of Abraham and Rendell. It is also worth pointing out that, by this measure, there is room for the participation rate to continue to move up should the labor market continue to improve, in line with normal cyclical behavior.

As an alternative, we could do an exercise more similar to the one that Abraham and Rendell did, where we compute the change in the participation rate and subtract out the part we can explain—the change in the predicted

Figure 2. Labor Force Participation Rate: Actual, Trend, and Predicted, 1990–2022



Sources: Author's calculations using data from the Current Population Survey, the US Census Bureau, the Bureau of Labor Statistics, the Congressional Budget Office, and Aaronson and others (2014).

Note: The labor force participation rate is calculated from the Current Population Survey using weights, provided by John Coglianese and Christopher Nekarda, that reflect the US Census Bureau's Vintage 2021 population estimates. The data were seasonally adjusted using factors from the Bureau of Labor Statistics (actual and predicted labor force participation rate). The trends have also been adjusted to account for the Vintage 2021 population estimates.

value. In this case, the unexplained portion of the decline is about 0.5 percentage point.[3]

As noted above, the actual participation rate is in line with its predicted value, but the change in the actual participation rate relative to the change in the predicted value suggests that about 0.5 percentage point of the decline in the participation rate is unexplained. What explains the difference? The second calculation captures the fact that the participation rate was about 0.5 percentage point above the predicted value just prior to the pandemic.

Based on this accounting, what you expect to see happen to the participation rate going forward depends on why you think the participation rate was so far above the predicted value prior to the pandemic and whether it is likely to persist. For instance, if the high level of the participation rate was just noise, or if we underestimated the cyclical position of the economy at that time, there is no reason to think that the high level of participation from that time would persist going forward. If, however, the unusually high level reflected a higher pre-pandemic trend, then we might expect a larger recovery.

3. In comparison the most comparable estimate of the unexplained gap in Abraham and Rendell is 0.3 percentage point. Of this difference, 0.1 percentage point can be explained by the fact that I don't adjust for the 2022 Vintage population controls and 0.1 percentage point is accounted for by the fact that I am looking at quarterly data.

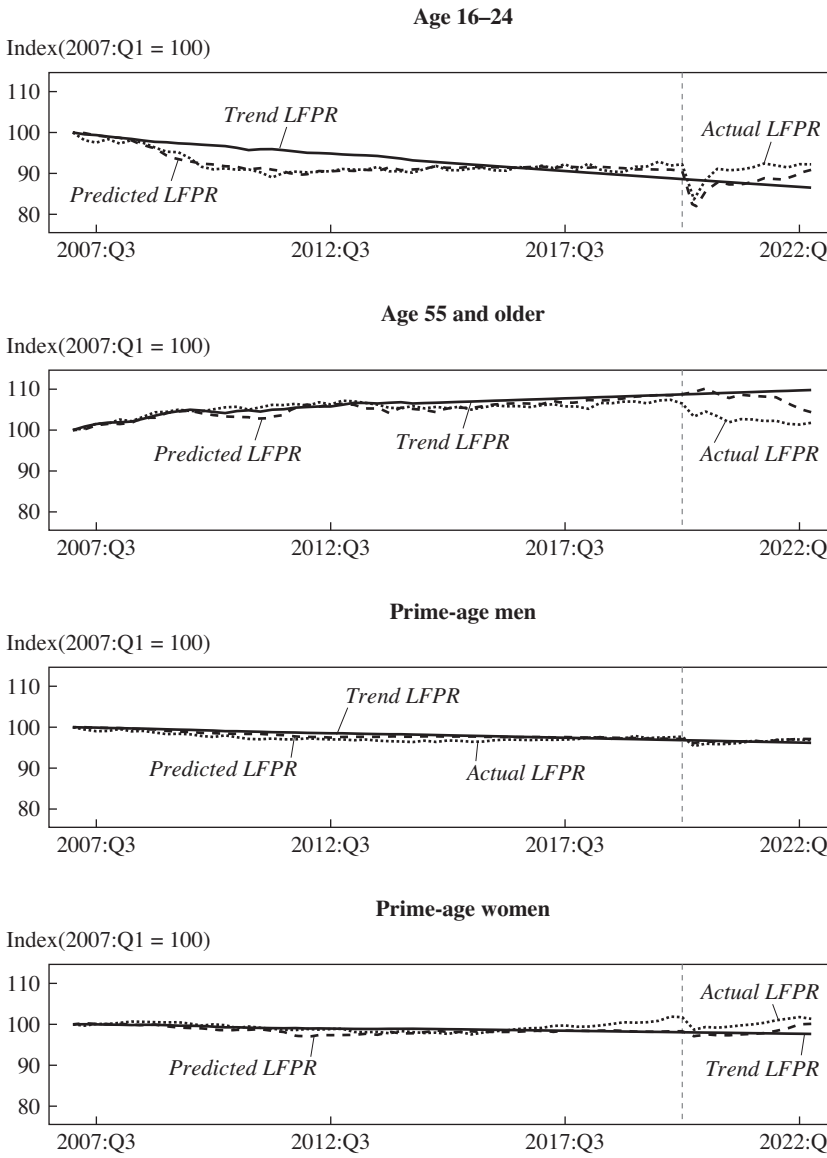
To elucidate whether there might be some signal for the trend in the relatively high pre-pandemic participation rate, figure 3 shows the participation rates of a few demographic groups along with their trends from the 2014 paper. The participation rates were calculated using the Current Population Survey (CPS) data and weights that account for the Vintage 2021 population controls. The data were seasonally adjusted using factors from the BLS. Note that I haven't adjusted the trends for the new population controls.[4] However, since the trends in the paper stopped in 2014, I drew out the trends using the growth rates over the 2009–2014 period. I also estimated the cyclical behavior of participation for each of the groups and plotted the predicted values. To put everything on the same scale, the data are indexed to 2007:Q1. A vertical dashed line marks the start of the pandemic.

These graphs show that for some groups there are differences in the evolution of their participation rates relative to their trends in recent years, which might have some information for thinking about the participation rate going forward. To start, the top graph shows the participation rate among those 16–24 years old. The participation rate for young people flattened out starting in 2012 and has largely moved sideways since then. The overall flattening is in line with what would be expected given the cyclical recovery in the economy over the time (comparing the actual participation rate (the dotted line) to the fitted value (the dashed line). That said, in the year just prior to the pandemic, and even more so now, the participation rate stands above what you would expect given the cyclical position of the economy. While our alternative trend captures this flattening better than the baseline (which is represented in this figure), nonetheless the relative strength in youth participation could suggest that the trend participation rate is higher than that embodied by even our alternative model.

Participation among prime-age women (the bottom graph) is also high relative to its trend and relative to what we would expect given the state of the business cycle. As with young people, this behavior started prior to the pandemic. The participation rate for prime-age women rose for much of the post–World War II era, up until about the year 2000, at which point it flattened out and even started to edge down. Since much of the weakness over the past two decades was among low-skilled women (Black, Schanzenbach, and Breitwieser 2017), it raised the question of whether their participation rate was being dragged down by the same demand factors

4. Because we divide the workers into broad age categories, this partially mitigates the fact that the trends are not adjusted for the new population controls.

Figure 3. Labor Force Participation Rates for Selected Age Groups



Sources: Author's calculations using data from the Current Population Survey, the US Census Bureau, the Bureau of Labor Statistics, the Congressional Budget Office, and Aaronson and others (2014).

Note: The labor force participation rate is calculated from the Current Population Survey using weights, provided by John Coglianese and Christopher Nekarda, that reflect the US Census Bureau's Vintage 2021 population estimates. The data were seasonally adjusted using factors from the Bureau of Labor Statistics (actual and predicted labor force participation rate). The trends have also been adjusted to account for the Vintage 2021 population estimates. Vertical dashed line indicates 2020:Q1.

dragging down the participation of low-skilled men. But in the years just prior to the pandemic and now during the recovery, the participation rate among prime-age women has started to edge up and now stands at about its historical peak. This may in part be due to the substantial increase in female educational attainment in recent years.[5] The more recent movements could also be due to changes in the labor market post-pandemic, including the persistence of remote work. While our model captures changes in educational attainment, it does so in a blunt way, and the model doesn't capture remote work at all. If these factors persist in boosting female labor force participation, then the actual trend participation rate is higher and flatter than that shown in figure 1.

In contrast, the participation rate of prime-age men appears to have continued to fall, roughly in line with the trend. It was perhaps a touch above trend prior to the pandemic and now, but it isn't clear that there has been a break relative to the model expectations.

The clearest break relative to model expectations seems to be among workers 55 and older, whose participation had been rising roughly in line with trend but fell off right as the pandemic hit. This makes the point clearly that, to the extent that there is a shortfall in participation, it is among retired workers (Montes, Smith, and Dajon 2022). The implications for labor force participation moving forward depend on whether the spate of early retirements was a onetime response to the pandemic. If so, the participation rate will return to its trend as these people pass the age at which they would have retired under normal circumstances. If, instead, COVID-19 changed workers' preferences about retirement—either because it has increased the health risks to working while older, or because it has affected the preference for work more generally, then this could represent a lower trend participation rate going forward, suggesting that the actual participation is even closer to its trend than the current model suggests.

To sum up this discussion of the LFPR, the evidence presented by Abraham and Rendell and in the alternative benchmarking provided by the model suggests that the LFPR is not particularly low. I would estimate the shortfall in the participation rate to be around zero, in line with the reading from the model. It is possible that the participation rate is further below its

5. Institute of Education Sciences National Center for Education Statistics, "Digest of Education Statistics 2021," Table 104.10 Rates of high school completion and bachelor's degree attainment among persons age 25 and over, by race/ethnicity and sex: Selected years, 1910 through 2021.

expected level, if the labor market is tighter than suggested by the unemployment rate gap, or if the trend is higher, for instance, because the trend in participation among women is higher than the model embodies. On the other hand, participation could actually be closer to the trend, if the change in retirement behavior during the pandemic is persistent, suggesting the trend is lower than that embodied in the model.

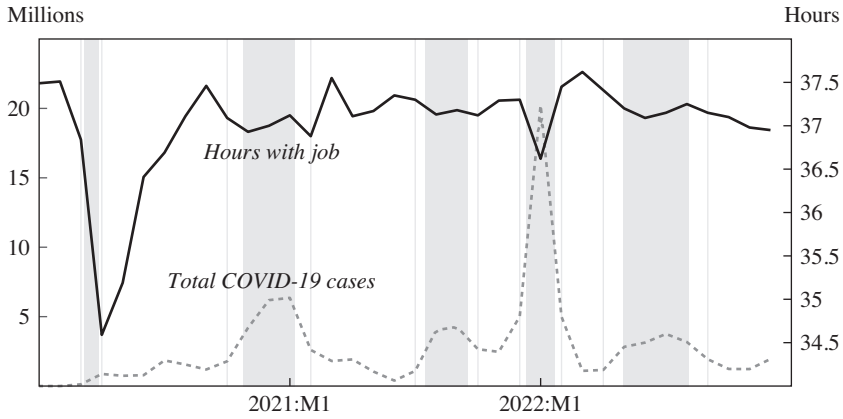
Up to now I have focused my discussion on the LFPR. However, the authors parse the decline in labor supply, as measured by total hours, between the participation rate and average weekly hours of work. By their calculation, when accounting for preexisting trends, the workweek accounts for perhaps three times the unexplained reduction in total hours as does the LFPR. Unfortunately, the authors have much less to say about the change in hours than they do participation. In part this is because the studies they review tended to focus on employment or participation and not hours. Given its importance to recent labor supply trends, the paper clearly identifies the workweek as an area requiring more attention by researchers. I will try to provide some guidance for future work on the topic.

The authors and a number of the studies they discuss (Goda and Soltas 2023; Sheiner and Salwati 2022) focus on the impact of COVID-19 on the level of hours. On aggregate, these studies and the additional work done by Abraham and Rendell find evidence that there has been an increased rate of illness since the pandemic, likely related to bouts of COVID-19 and especially long COVID, which has affected the level of hours worked. However, this work misses the fact that the pandemic has also ushered in a period of increased volatility in hours.

Figure 4 shows the Abraham and Rendell measure of hours along with periods of spikes in COVID-19 cases. As can be seen, the two are negatively correlated, with weekly hours of work falling when COVID-19 cases rise dramatically. More generally, I calculated the variance of hours before the pandemic and from fall 2021, when schools really reopened, through the end of 2022, and it is about 50 percent higher. The higher volatility could be due to any number of factors, including changes in hours due to sickness or to caregiving. And since higher volatility has its own cost associated with it—for instance, by reducing the ability of firms to plan their labor inputs and making it harder for individuals to balance work and family care—this is a phenomenon worth exploring.

Moving on from the COVID-19 effects, the authors put out the tantalizing possibility that perhaps people are changing their preferences around how many hours they work. This is an interesting idea, but I don't think that the measure of hours that the authors provide is particularly suggestive

Figure 4. Hours Worked Last Week and COVID-19 Cases



Sources: Author's calculations using data from the US Census Bureau, the Bureau of Labor Statistics, and the *New York Times*.

Note: Hours worked last week include zero for individuals who reported being with a job but not at work. Hours data were seasonally adjusted using the Census Bureau's X-13 method. Shading shows approximate dates of COVID-19 waves. COVID-19 case data retrieved from https://github.com/nytimes/covid-19-data using the variable cases_avg.

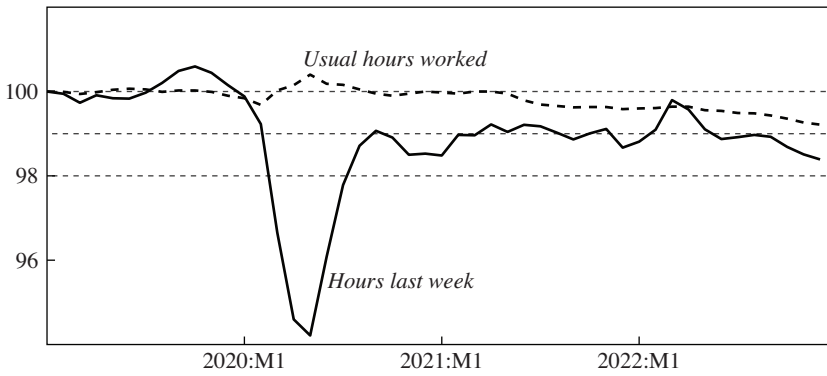
of this phenomenon, since it captures actual hours of work rather than desired hours of work. Actual hours of work can change when, as was just discussed, someone gets sick or if they have caregiving responsibilities. It can also vary due to weather, if employers adjust hours of work in response to changes in demand, and in response to other temporary changes in personal circumstances.

Luckily for us, the CPS also includes a measure of usual hours worked, which is closer conceptually to desired hours of work. Of course, there is a large literature suggesting people don't have complete choice when it comes to the hours they can work.[6] But at least the usual hours of work variable is free from the week-to-week shocks that move people away from their desired hours. In addition, my measure differs from the usual weekly hours measure published by the BLS because I eliminate from my sample people who are working part-time for economic reasons, since they are clearly off their supply curves.

Figure 5 compares the authors' measure of hours worked last week with my version of usual hours worked. To make the trend easier to see, I indexed

6. See Altonji and Paxson (1988) for a relatively early empirical exploration and Golden and Gebreselassie (2007) for a more recent example.

Figure 5. Alternative Hours Measures



Sources: Author's calculations using data from the US Census Bureau and the Bureau of Labor Statistics.

Note: Hours worked last week include zero for individuals who reported being with a job but not at work. Usual hours worked excludes those who worked part-time for economic reasons. Hours data were seasonally adjusted using the Census Bureau's X-13 method. Both series are centered three-month moving averages.

both series to be equal to 100 in 2019:M1. After falling dramatically during the pandemic, the authors' measure rebounded to a level about 1.25 percent below its pre-pandemic level and has moved about sideways since then. This timing is supportive of the authors' hypothesis that pandemic-related factors, such as fear of COVID-19 and increased illness, can explain the decline in actual hours worked. In contrast, usual hours worked was flat on balance, until the middle of 2021—suggesting that desired hours hadn't changed, even as actual hours declined. However, since mid-2021, around the time vaccines became more prevalent, usual hours of work has been trending down. I think that this provides some evidence that, as people have adjusted to post-pandemic life, desired hours may have fallen. If that is true, it would have long-term implications for potential output.

A final point I wish to make with respect to identifying missing workers is that by ignoring population, the authors miss out on an important part of the story for why the labor market has been so tight. Abraham and Rendell acknowledge in a footnote that they focus on people's willingness to supply their labor—through the LFPR and weekly hours—and ignore the population. But I think it is worth recalling just how dramatic the downshift in population growth was during the pandemic, and even prior to that. The growth rate of the working-age population fell from an average of about

1.25 percent between 1980 and 2007 to about 1 percent between 2007 and 2019. During the pandemic population growth slowed to about half that. Excess mortality during the pandemic alone accounted for about 500,000 fewer people in the working-age population (Federal Reserve System Board of Governors 2023). In addition, the shortfall in immigration between 2020 and 2022 amounts to an additional 750,000 people.[7] If immigrants had the same participation rates as the working-age population, this would have implied a shortfall of about 460,000 labor force participants in 2022:Q4. However, immigrants have much higher participation rates, so the number of missing participants due to immigration is probably closer to the upper bound. Altogether excess deaths and immigration probably account for between 1 and 1.25 million missing labor force participants. In comparison, our estimates suggest that the *unexpected* shortfall in the LFPR in the fourth quarter of 2022 was somewhere between zero and 0.5 percentage point. If we apply this miss to the working-age population in 2022:Q4, we find that the participation rate accounts for somewhere between zero and 1.3 million missing workers. By this accounting, reductions in the population account for somewhere between all of the missing workers and half of them.[8]

In terms of explaining the impact of the pandemic on labor supply, the authors left several areas underexplored, which point to areas for future research. Abraham and Rendell note that the COVID-19-related factors overexplain the decline in LFPR, depending on the exact calibration, but at the same time they cannot explain changes in labor supply for different demographic groups. As discussed here, in their paper, and elsewhere, the real shortfall in participation is among older people, and the work on long COVID cannot explain the magnitude of the change for that group. At the same time, if the aggregate estimates of the impact of COVID-19 are in the right ballpark, this suggests that there have been offsetting factors increasing participation among other groups. Determining which workers have increased their participation and why is important for understanding their implications for potential labor supply in the future.

7. In 2018 and 2019 immigration averaged about 950,000 people. In 2020–2022, immigration averaged about 700,000 people. If we assume that, in the absence of the pandemic, immigration would have remained at the 2018–2019 pace, then the shortfall in immigrants over the three years is about 750,000. For source data, see Knapp and Lu (2022).

8. These findings are similar to those by Bauer and others (2023), who put the labor force shortfall at 900,000, all of which they attribute to excess deaths and reduced immigration.

I also believe the authors were too dismissive of the possibility that the large expansion of the social safety net had an impact on labor supply. Perhaps due to space or time constraints, they did not delve into the details of the various income support programs. More generally, it is not enough to focus on the average benefit per workers—the distribution of benefits matters substantially for their likely impact on labor supply, since households with different levels of income face very different budget constraints. Moreover, it is not clear that the income and substitution effects estimated in a non-pandemic environment are useful for assessing likely behavior during a pandemic, when people are confronted with significant health risks and reductions in the availability of child and family care among other changes. While it is true that the social safety net programs have expired, they nonetheless may have had an impact on people's work habits that could be persistent.

While acknowledging that no paper can be comprehensive, I was a bit surprised that the authors didn't tackle the issue of remote work, the widespread use of which at the height of the pandemic transformed many jobs and people's expectations of their work and home life. This transformation could have important implications for both labor force participation and the workweek going forward. For instance, as described by Bick, Blandin, and Fuchs-Schündeln (2022), to the extent that remote work lowers the fixed costs of work, for instance, by reducing commute times and allowing for greater flexibility in location and family care, it could increase labor force participation and reduce hours of work. The extent to which remote work will continue to be broadly available to workers whose jobs can feasibly be done remotely will have important implications for the size and composition of the labor force going forward.

To conclude, the authors have provided a wide-ranging first look at the evolution of the labor market as we emerge from the pandemic. The work included an accounting of the role of the LFPR and the workweek in contributing to the shortfall as well as a careful appraisal of a number of pandemic-related factors that could be affecting labor supply. Their work also points to several important areas that require future research.

With respect to the implications for the economy, the research done by the authors and in this discussion suggest that most of the shortfall in labor force participation relative to its pre-pandemic level has largely been as expected. This implies that further expansion of labor supply will depend on continued cyclical improvement in the participation rate and the workweek and a rebound in immigration that is apparently already under way. Another potential source of labor supply would involve the reversal of some

persistent pandemic-induced patterns, including both early retirements and the low level of hours worked, or further positive structural changes, for instance, from remote work or changing demographics, such as higher educational attainment. Policymakers will certainly be paying attention.

REFERENCES FOR THE AARONSON COMMENT

Aaronson, Stephanie, Tomaz Cajner, Bruce Fallick, Felix Galbis-Reig, Christopher Smith, and William L. 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 F. Pingle, and William L. 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: 69–154.

Altonji, Joseph G., and Christina H. Paxson. 1988. "Labor Supply Preferences, Hours Constraints, and Hours-Wage Trade-offs." *Journal of Labor Economics* 6, no. 2: 254–76.

Bauer, Lauren, Wendy Edelberg, Sara Estep, and Brad Hershbein. 2023. "Who's Missing from the Post-Pandemic Labor Force?" Washington: Brookings Institution. https://www.brookings.edu/2023/04/04/whos-missing-from-the-post-pandemic-labor-force/.

Bick, Alexander, Adam Blandin, and Nicola Fuchs-Schündeln. 2022. "Reassessing Economic Constraints: Maximum Employment or Maximum Hours?" In *Economic Policy Symposium Proceedings: Reassessing Constraints on the Economy and Policy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas.

Black, Sandra E., Diane Whitmore Schanzenbach, and Audrey Breitwieser. 2017. "The Recent Decline in Women's Labor Force Participation." Washington: The Hamilton Project, Brookings Institution. https://www.brookings.edu/wp-content/uploads/2017/10/es_10192017_decline_womens_labor_force_participation_blackschanzenbach.pdf.

Federal Reserve System Board of Governors. 2023. *Monetary Policy Report—June 2023*. Washington: Author. https://www.federalreserve.gov/monetarypolicy/2023-06-mpr-summary.htm#:~:text=Monetary%20Policy%20Report%20submitted%20to,FOMC)%20objective%20of%202%20percent.

Goda, Gopi Shah, and Evan J. Soltas. 2023. "The Impacts of Covid-19 Absences on Workers." *Journal of Public Economics* 222:104889.

Golden, Lonnie, and Tesfayi Gebreselassie. 2007. "Overemployment Mismatches: The Preference for Fewer Work Hours." *Monthly Labor Review*, April, 18–37.

Knapp, Anthony, and Tiangeng Lu. 2022. "Net Migration between the United States and Abroad in 2022 Reaches Highest Level since 2017." US Census Bureau, December 22. https://www.census.gov/library/stories/2022/12/net-international-migration-returns-to-pre-pandemic-levels.html.

Montes, Joshua, Christopher Smith, and Juliana Dajon. 2022. "'The Great Retirement Boom': The Pandemic-Era Surge in Retirements and Implications for Future Labor Force Participation." Finance and Economics Discussion Series 2022-081. Washington: Board of Governors of the Federal Reserve System. https://doi.org/10.17016/FEDS.2022.081.

Sheiner, Louise, and Nasiha Salwati. 2022. "How Much Is Long COVID Reducing Labor Force Participation? Not Much (So Far)."Working Paper 80. Washington: Hutchins Center on Fiscal and Monetary Policy, Brookings Institution. https://www.brookings.edu/research/how-much-is-long-covid-reducing-labor-force-participation-not-much-so-far/.

COMMENT BY

YONGSEOK SHIN Abraham and Rendell's superb article provides a thoughtful and thorough analysis of the trends in labor force participation and hours of work since the pandemic. Much has been written about the labor shortage during the post-pandemic recovery. Abraham and Rendell convincingly show that the lower participation rate is a continuation of a trend that existed even before the pandemic. In addition, the little shortfall that remains after adjusting for the trend can be attributed to people withdrawing from the labor force because of long COVID or fear of COVID-19, although the magnitude of these effects is smaller than what other researchers have posited. As for the decline in hours of work, Abraham and Rendell note that there was no preexisting trend, and that long COVID or fear of COVID-19 may not be the explanation, to the extent that these effects will predominantly operate at the extensive margin (working or not working) rather than at the intensive margin (hours of work, conditioning on working). The authors conjecture that the lower levels of hours reflect a reevaluation of the work-life balance people wish to strike.

I will first comment on the trend in labor force participation rates and offer some corroborating evidence on the role of fear of COVID-19. I will then discuss how the abrupt reallocation of economic activities during the pandemic is another reason for the tight labor market. Finally, I will point to the patterns in the declining hours of work across workers, which support the authors' conjecture on the cause of reduced hours of work.

TREND IN LABOR FORCE PARTICIPATION The labor force participation rate in the United States has been declining since 2000, as shown in figure 2 of the paper. This is partly due to demographic changes (population aging), but also due to downward trends within demographic groups. An alternative view of figure 2 is that it may not be one monotonic trend. For example, between 2004 and 2008, the participation rate stayed flat until hit

by the Great Recession. The downward trend was again arrested in 2014, with the participation rate increasing slightly through 2019 until hit by the pandemic lockdown. While the compositional effect of the demographic changes may well be predetermined and monotonic, the participation rates within demographic groups may have behaved differently had it not been for the Great Recession or the pandemic. In other words, we may not want to take all the preexisting downward trend in the participation rate as inexorably given.

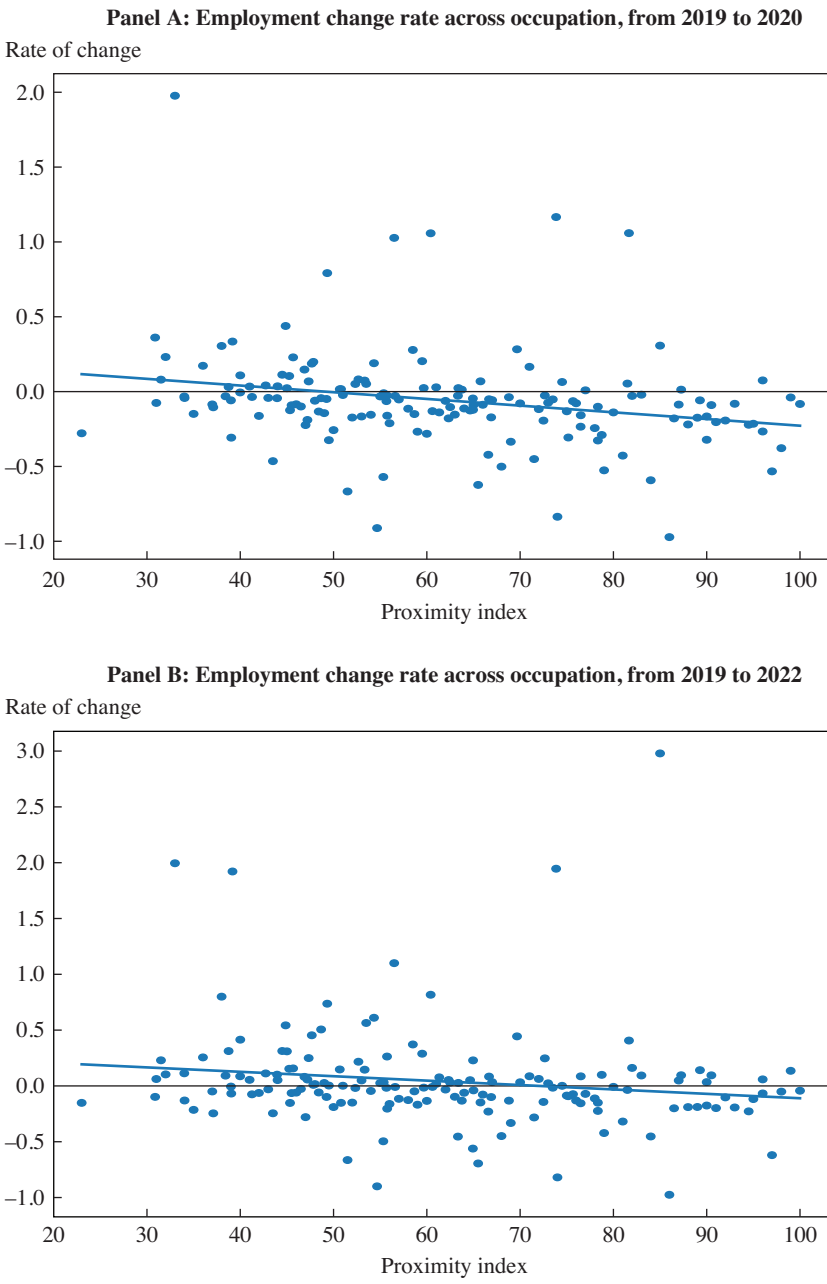
The real question then is why the participation rates of some demographic groups are trending downward. As is well known, the labor force participation rate of young men without a four-year college degree has been falling since the Great Recession. Lee, Park, and Shin (2023) show that the participation rate of women over 30 without a four-year college degree has been falling as well. Considering the personal and social losses that young people's detachment from the labor force entails, it is imperative that researchers and policymakers delve into the cause of this phenomenon and come up with policy responses.

FEAR OF COVID-19 The discussion in the paper on the effect of fear of COVID-19 largely relied on direct survey evidence. I provide two pieces of corroborating evidence, building on my previous work on this topic.

First, I consider the impact of fear of COVID-19 on occupational employment. The idea is that occupations differ in terms of COVID-19 infection risks, and employment in riskier occupations would shrink more if fear of COVID-19 were an important factor.

In figure 1, panel A plots the change in employment of occupations between 2019 and 2020 against each occupation's proximity index. The proximity index is a weighted index of job attributes measuring the extent to which the job requires workers to perform tasks in close physical proximity of other people (Aum, Lee, and Shin 2022). Given how COVID-19 gets transmitted, proximity at work is a good proxy for the risk of COVID-19 infection at work. We see that employment in occupations with high COVID-19 infection risks fell relative to those with low risks, suggesting that fear of COVID-19 had an impact on the labor market in 2020. Panel B shows the change in employment of occupations between 2019 and 2022. Although the pandemic had receded significantly by 2022, we still observe a significant negative relationship between the measure of COVID-19 infection risks at work and occupational employment. This is consistent with Abraham and Rendell's finding that fear of COVID-19 negatively affected people's labor market activities even in late 2022.

Figure 1. Occupational Proximity at Work and Employment Change



Sources: Author's calculations using the Current Population Survey and the occupational proximity index of Aum, Lee, and Shin (2022).

Table 1. Regressing Nonemployment Indicator on State-Level COVID-19 Cases

	Medium-risk state	*High-risk state*
Estimated coefficient	0.007	0.013
Standard error	0.001	0.001

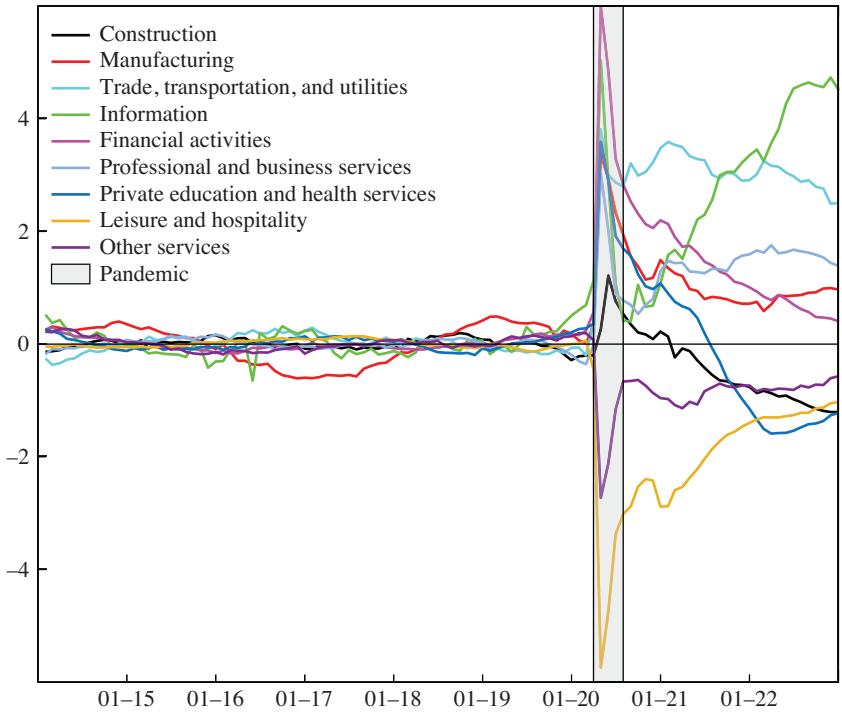
Sources: Author's calculations using the Current Population Survey and the Centers for Disease Control and Prevention, "COVID Data Tracker," https://covid.cdc.gov/covid-data-tracker/#datatracker-home.

Second, I explore whether concurrent COVID-19 cases have a negative effect on employment, with the idea that fear of COVID-19 will be heightened when and where the number of new COVID-19 cases is large. I extend the work of Lee, Park, and Shin (2021) and run an individual-level regression pooling the Current Population Survey (CPS) for all twelve months of 2022. The dependent variable is an indicator for nonemployment (nonparticipation and unemployment). The control variables include saturated dummy variables for individual characteristics such as gender, race or ethnicity, age, and education. In addition, I group US states into high, medium, and low risk in each month based on the number of new COVID-19 cases per 1,000 people during the four weeks that precede the week of the CPS interview (the week that contains the 19th of each month), using case counts from the Centers for Disease Control and Prevention.

In table 1, we see that residing in a state classified as medium COVID-19 risk at the time of the CPS interview is associated with a 0.7 percentage point increase in nonemployment rate, relative to the (excluded) low-risk states. People in states classified as high COVID-19 risk at the time of the CPS interview have a nonemployment rate that is 1.3 percentage points higher than those in low-risk states. As it is reasonable to think that a large number of recent COVID-19 cases will strengthen fear of COVID-19, the estimated coefficients corroborate Abraham and Rendell's finding that fear of COVID-19 continued to be a drag on labor force participation and employment through 2022.

PANDEMIC AS A REALLOCATION SHOCK The pandemic lockdown was not just a pause of the economy. It also reallocated economic activities across sectors, as emphasized by Buera and others (2021). In figure 2, I plot the employment shares of two-digit sectors of the US economy from January 2014 to January 2023. For each sector, I extrapolate a linear trend using the pre-pandemic data from 2014 and 2019 and then plot the deviation of the employment share from the trend, normalized by the standard deviation calculated from the pre-pandemic data.

Figure 2. Employment Share of Two-Digit Industries, Detrended and Standardized



Source: Author's calculations using the Current Employment Statistics survey.

Figure 2 shows that the pandemic had an unprecedentedly large effect on employment shares of sectors and that it affected sectors in a starkly asymmetrical fashion. What is remarkable is that the sectoral employment shares are still far away from the pre-pandemic trends a full three years after the initial shock. As worker skills are partly sector-specific, the pandemic as reallocation shock can generate a mismatch between available jobs and workers, which can explain both the exceptional tightness of the current labor market and the low labor force participation. Workers displaced from shrinking sectors cannot easily move to expanding ones because of skill mismatch, which may lead them to exit the labor market altogether. This is a channel that deserves more scrutiny.

DECLINE IN HOURS OF WORK The total work hours of an economy can fall because either the number of workers falls or workers work fewer hours.

As Abraham and Rendell note, a decomposition shows that the decline in total work hours between 2019 and 2022 is as much due to fewer hours per worker as fewer workers (Lee, Park, and Shin 2023). The hours margin becomes more important when extending the analysis to early 2023 because the participation rate continued to rise between January and April 2023. The participation rate in April 2023 is only 0.37 percentage points below the average participation rate between 2017 and 2019, without adjusting for any preexisting trend. On the other hand, hours per worker did not increase in 2023, as of this writing.

Between 2019 and 2022, the average annual work hours per person in the United States fell by 36 hours from 1,229 to 1,193, or by 3 percent. Of these, 19 hours are due to the hours margin and the remaining 17 are due to the number of workers margin (Lee, Park, and Shin 2023).

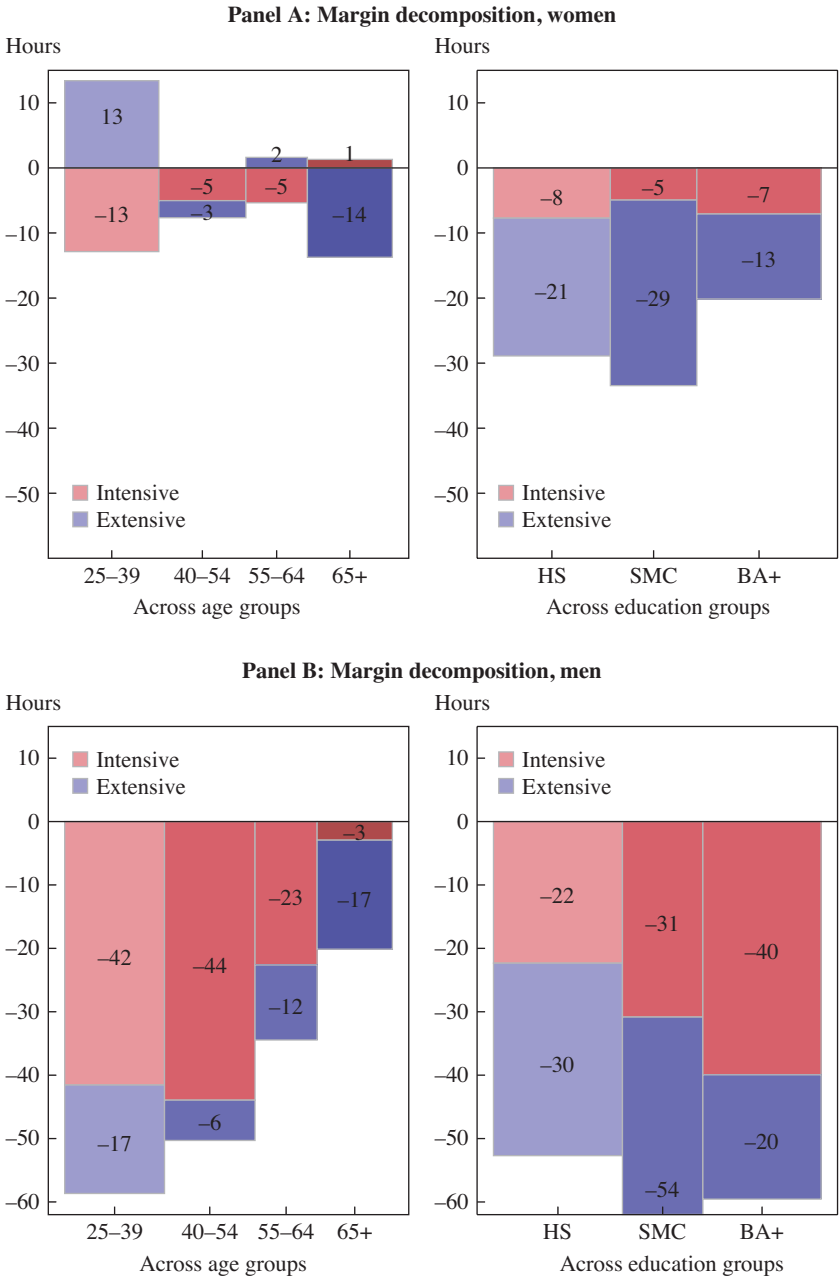
In figure 3, I divide the population by gender, age, and education to decompose the annual work hours changes into the hours margin (intensive) and the number of workers margin (extensive) for each demographic group. Panel A is for women and panel B is for men. For both women and men, the bars on the left represent age groups: 25–39, 40–54, 55–64, and 65 and older. The bars on the right are groups by educational attainment: high school graduates and dropouts (HS), those with some college education but no four-year degree (SMC), and those with a four-year degree or even more education (BA+).

Figure 3 clearly shows that the hours reduction between 2019 and 2022 was much larger for men than women. Men's annual work hours per person fell by 60 hours, while women's fell by 15.[1] And for men, the hours margin was the dominant one for all age groups except for those 65 and older. By education, the hours margin was especially important for the most educated group. Overall, the work hours per worker (that is, excluding those not working) fell the most among educated, prime-age men, who tend to work long hours and have high earnings. This last finding, together with the excess demand in the overall labor market, suggests that the reduction in hours is a voluntary decision, possibly because of a desire to strike a better work-life balance, as Abraham and Rendell conjectured.

CONCLUDING REMARKS Abraham and Rendell's masterful review convincingly accounts for the decline in labor force participation through the end of 2022. It also raises a few thought-provoking questions. Why does

1. Adding up the bars using population weights does not necessarily yield the averages for women and men because the bars do not capture the changes in the population weights of the age or education groups.

Figure 3. Change in Annual Work Hours per Person, by Gender, Age, and Education



Sources: Author's calculations using the Current Population Survey and the decomposition method in Lee, Park, and Shin (2023).

the labor force participation of young men and women without college degrees continue to fall? What explains the historic tightness of the current labor market? Did the massive reallocation of economic activities across sectors play a role? Most importantly, what are the reasons that many educated, prime-age men reduced their work hours, and is this a permanent change? I expect that future research by Abraham and Rendell, and by other researchers inspired by their work, will provide insights into these important questions.

REFERENCES FOR THE SHIN COMMENT

Aum, Sangmin, Sang Yoon Lee, and Yongseok Shin. 2022. "Who Should Work from Home during a Pandemic? The Wage-Infection Trade-off." *Federal Reserve Bank of St. Louis Review* 104, no. 2: 92–109. https://research.stlouisfed.org/publications/review/2022/01/24/who-should-work-from-home-during-a-pandemic-the-wage-infection-trade-off.

Buera, Francisco J., Roberto N. Fattal-Jaef, Hugo Hopenhayn, P. Andres Neumeyer, and Yongseok Shin. 2021. "The Economic Ripple Effects of COVID-19." Working Paper 28704. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28704.

Lee, Dain, Jinhyeok Park, and Yongseok Shin. 2023. "Where Are the Workers? From Great Resignation to Quiet Quitting." Working Paper 30833. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30833.

Lee, Sang Yoon, Minsung Park, and Yongseok Shin. 2021. "Hit Harder, Recover Slower? Unequal Employment Effects of the COVID-19 Shock." *Federal Reserve Bank of St. Louis Review* 103, no. 4: 367–83. https://research.stlouisfed.org/publications/review/2021/09/01/hit-harder-recover-slower-unequal-employment-effects-of-the-covid-19-shock.

GENERAL DISCUSSION Caroline Hoxby commented on a work-life balance evaluation that is ongoing, especially during the COVID-19 pandemic. For years, the United States has been an outlier in terms of hours and weeks of participation in a year. In the Scandinavian and continental European countries, people work fewer hours.[1] It's possible that the pandemic shocked Americans into thinking they didn't need to contribute so many hours or weeks of participation. So the question to ask is whether we're just converging to the norm or is there something else really going on.

1. OECD, "Average Annual Hours Actually Worked per Worker," https://stats.oecd.org/Index.aspx?DataSetCode=ANHRS.#

Jason Furman added that if we were looking only at labor market data from 2019 and before and now in 2023, there likely wouldn't be a paper or conversation about missing workers. He noted that in January 2020, Phillip Swagel, Wendy Edelberg, and Jeffrey Kling's forecast for the labor force participation rate in 2023:Q1 was 62.5 percent and that in February 2023, the labor force participation rate was 62.5 percent.[2] Employment is exceeding what Congressional Budget Office (CBO) had forecasted prior to the pandemic, which is remarkable when considering that the population, due to premature deaths and immigration, is smaller than what the CBO had forecast.[3] A 90 percent confidence interval around these forecasts in 2019 would have accurately predicted where we are today, and we would have had instead a paper about why employment is so surprisingly high right now.

Furman also questioned the paper's dismissal of persistent effects of labor supply programs, arguing that a vigorous recession response can be partly explained by the worry about unemployment in one year translating to future unemployment as well, such that even if the benefits are gone, we don't know if it's just a wealth effect. He also called attention to the difference in response between the United States and Europe, where Europe had a better employment response and less of an hours response.

Steven Davis spoke about the correct way to think about the shortfall in the labor force participation rate. Stephanie Aaronson and Katharine Abraham had both made the point that the new population controls introduced in 2022 in the Current Population Survey (CPS) caused the actual reduction to be understated by 0.3 percentage points.[4] He asked why that wasn't part of what we are trying to explain.

Also, in response to Furman's comment, Davis noted that the shift to remote work expanded labor market opportunities for several types of workers, who might be facing constraints such as mobility impairments or family obligations, or who are located in left-behind areas. The shift to remote work means an expansion of labor market opportunities and the

2. CBO, *The Budget and Economic Outlook: 2020 to 2030* (Washington: Congressional Budget Office, 2020), https://www.cbo.gov/system/files/2020-01/56020-CBO-Outlook.pdf; US Bureau of Labor Statistics, "Civilian Labor Force Participation Rate," https://www.bls.gov/charts/employment-situation/civilian-labor-force-participation-rate.htm.

3. Centers for Disease Control and Prevention, "COVID Data Tracker," https://covid.cdc.gov/covid-data-tracker/#demographics, https://www.census.gov/library/stories/2021/12/net-international-migration-at-lowest-levels-in-decades.html.

4. US Bureau of Labor Statistics, "Adjustments to Household Survey Population Estimates in January 2022," https://www.bls.gov/cps/population-control-adjustments-2022.pdf.

potential to draw some of these people into the labor force and push the labor force shortfall in the other direction.[5]

Davis then turned to fear of COVID-19 as an explanation in the paper for missing workers. In the summary table, the estimate is 0.3 percentage points but in figure 5, the average value from data in 2022 looks to be around 0.7 or 0.8 percentage points. He acknowledged that, in her presentation, Abraham explained there could be a sample selection or social desirability bias, which are reasonable concerns, but he suggested that that doesn't explain the pattern in figure 5, which is that the fear of COVID-19 is very large initially but declines by 2022.

Addressing Abraham's comments about measuring the fear of COVID-19, Davis added that an upcoming paper will have different numbers associated with fear of COVID-19 and other infection risks, smaller but still well above those in the Household Pulse Survey.[6] Lastly, he mentioned that, from past experience in surveying workers, there is a tendency for workers to include commuting time in hours worked. Yongseok Shin stressed in his discussion that the hours shortfall is showing up in the same groups that have cut back on their commute times. Davis questioned how much of the hours shortfall is a measurement issue, whereby the groups that say they are working less might simply be spending less time commuting.

Robert Hall agreed with Aaronson's comment that it's a mistake to identify the labor force participation rate—the sum of employment and unemployment—as a single variable that measures labor supply. The paper does break labor force participation into its two components by showing that, for some months, the unemployment rate before and after the pandemic was the same 3.5 percent, which is useful to untangle the separate

5. Adam Ozimek, "Remote Work Is Enabling Higher Employment among Disabled Workers," Economic Innovation Group, October 15, 2022, https://eig.org/remote-work-is-enabling-higher-employment-among-disabled-workers/; Arlene S. Kanter, "Our New Remote Workplace Culture Creates Opportunities for Disabled Employees," Harvard Law School, Petrie-Flom Center, Bill of Health, March 10, 2022, https://blog.petrieflom.law.harvard.edu/2022/03/10/remote-work-disability-ada/; Jose Maria Barrero, Nicholas Bloom, and Steven J. Davis, "Why Working from Home Will Stick," working paper 28731 (Cambridge, Mass.: National Bureau of Economic Research, 2021), https://www.nber.org/papers/w28731; Yuting Chen, Patricia Cortés, Gizem Koşar, Jessica Pan, and Basit Zafar, "The Impact of COVID-19 on Workers' Expectations and Preferences for Remote Work," working paper 30941 (Cambridge, Mass.: National Bureau of Economic Research, 2023), https://www.nber.org/papers/w30941.

6. Jose Maria Barrero, Nicholas Bloom, and Steven J. Davis, "Long Social Distancing," working paper 30568 (Cambridge, Mass.: National Bureau of Economic Research, 2023), https://www.nber.org/papers/w30568; to be published in the *Journal of Labor Economics* 41, no. S1 (October 2023).

influences.[7] But overall, Hall suggests it would be better to present this paper as giving important information about unemployment, employment, and hours instead of saying it is about labor supply.

John Haltiwanger brought up the limitations of the CPS in measuring self-employment. There was some evidence that, pre-pandemic, the CPS was not keeping up with administrative data on tracking employment, and there's a growing gap between the two.[8] While the CPS showed a moderate rise in self-employment, a brand-new series—the Business Formation Statistics—saw an enormous surge in what the US Census calls likely employers and likely non-employers. The surge for likely non-employers has been larger: in January and February of 2023, more than 50 percent higher on a monthly basis in terms of new applications for likely non-employers relative to the same months in 2019.[9] Haltiwanger suggests that this can tell a different story, especially given that this activity is missing in the CPS, and that this might even be related to the discussions on work-life balance.

Melissa Kearney brought attention to the question of childcare as an explanation for trends in the labor force participation rate and was surprised Abraham had not mentioned it. Kearney and Furman investigated this aspect in 2021, and Claudia Goldin looked at it in 2022, and they didn't find any evidence of childcare being a factor, which upset some people.[10] She suggested it might be worth discussing in the paper because it has been such a prominent issue.

Jonathan Pingle pointed out that negative labor supply shocks are fairly persistent for those over the age of 65. Even conditional on Furman's point, the shortfall is pretty large. The paper's results show that most of the remaining shortfall in participation is among those over 65. If that is

7. FRED, "Unemployment Rate," https://fred.stlouisfed.org/series/UNRATE.

8. Katharine G. Abraham, John C. Haltiwanger, Claire Hou, Kristin Sandusky, and James R. Spletzer, "Reconciling Survey and Administrative Measures of Self-Employment," *Journal of Labor Economics* 39, no. 4 (2021): 825–60, https://www.journals.uchicago.edu/doi/abs/10.1086/712187.

9. US Census Bureau, "Business Formation Statistics," https://www.census.gov/econ/bfs/index.html.

10. Jason Furman, Melissa Schettini Kearney, and Wilson Powell, "The Role of Childcare Challenges in the US Jobs Market Recovery during the COVID-19 Pandemic," working paper 28934 (Cambridge, Mass.: National Bureau of Economic Research, 2021), https://www.nber.org/papers/w28934; Claudia Goldin, "Understanding the Economic Impact of COVID-19 on Women," working paper 29974 (Cambridge, Mass.: National Bureau of Economic Research, 2022), https://www.nber.org/system/files/working_papers/w29974/w29974.pdf.

the case, then the next question is how much of that is expected due to the shock and how much of that is due to ongoing effects of COVID-19. Pingle also commented on the lack of increased take-up of Social Security as particularly interesting given the labor supply shortfall for those older workers.[11] Lastly, he mentioned cohorts. Older age groups' participation should resume rising as the younger cohorts age into the older groups. With rising education and life cycle labor force attachment of women, he noted that going forward the labor force participation trends should be rising among the older age groups.

Edelberg mentioned a related upcoming piece from the Hamilton Project looking at the labor force participation rate by age, race or ethnicity, and sex (but not education). Edelberg and coauthors found there was no net effect on the labor force participation rate from within-group changes in propensity to work from 2019 to February 2023.[12] However, Abraham and Lea Rendell get 0.5 percentage points, which could be due to the comparison period being a bit different. In 2019, the labor force participation rate was a little bit lower than the twelve-month period through February 2020. The labor force participation rate in February 2023 was three-tenths higher than the 2022 average.[13] Using backcast data, they find that the reductions in population growth point to a smaller labor force by about a million people. In response to Furman's comment about CBO projections, Edelberg noted that if the population estimates that the CBO put out in January 2022 were available, there would have been a larger projected level of employment. Therefore, she suggested, it is not a good idea to take the level of employment they had projected and compare it to the current revised data.

Alan Blinder asked if there might be some information in the geographic differences across states in the incidence of COVID-19 and vaccination rates that could be helpful in attempting to gauge to what extent the changes

11. Social Security Administration, Annual Statistical Supplement, 2021, OASDI Benefits Awarded: Retired Workers, Table 6.B3, https://www.ssa.gov/policy/docs/statcomps/supplement/2021/6b.pdf; Owen Davis, "Employment and Retirement among Older Workers during the COVID-19 Pandemic," working paper 6 (New York: Schwartz Center for Economic Policy Analysis, the New School for Social Research, 2021), https://www.economicpolicyresearch.org/images/docs/research/nssr_working_papers/NSSR_WP_062021.pdf.

12. Lauren Bauer, Wendy Edelberg, Sara Estep, and Brad Hershbein, "Who's Missing from the Post-Pandemic Labor Force?" (Washington: Brookings Institution, 2023), https://www.brookings.edu/2023/04/04/whos-missing-from-the-post-pandemic-labor-force/.

13. US Bureau of Labor Statistics, "Civilian Labor Force Participation Rate," https://www.bls.gov/charts/employment-situation/civilian-labor-force-participation-rate.htm.

in labor supply are due to COVID-19 or something else, and then holding this up against where the jobs were missing or the labor force declined.

Abraham commented that there are legitimate questions about the right benchmark to use for evaluating current participation but concluded that no matter the benchmark used, there isn't a very large decline in participation left to explain. Responding to Aaronson's suggestion, she agreed that it could be informative to look at usual hours in addition to actual hours. She also expressed appreciation for Shin's comments about how declines in hours and participation are related to COVID-19 exposure, as well as how those declines vary across occupations and across states. His finding that it is the educated high earners who are working fewer hours is consistent, she suggested, with the idea that those who can afford it are driving a change in work-life balance.

In response to Kearney's question about childcare, Rendell confirmed that they did not find any notable differences in patterns between parents and nonparents during their analysis, in line with the findings of Furman and Kearney.[14] Regarding questions about retirement, a recent Federal Reserve paper found that these excess retirements represent individuals choosing to retire a little earlier than planned, but over the next few years the retired share of population should return to its pre-pandemic path.[15]

14. Furman, Kearney, and Powell, "The Role of Childcare Challenges."

15. Joshua Montes, Christopher Smith, and Juliana Dajon, "'The Great Retirement Boom': The Pandemic-Era Surge in Retirements and Implications for Future Labor Force Participation," Finance and Economics Discussion Series 2022-081 (Washington: Board of Governors of the Federal Reserve System, 2022), https://doi.org/10.17016/FEDS.2022.081.

JOHN E. T. BISTLINE
Electric Power Research Institute

NEIL R. MEHROTRA
Federal Reserve Bank of Minneapolis

CATHERINE WOLFRAM
Massachusetts Institute of Technology

Economic Implications of the Climate Provisions of the Inflation Reduction Act

ABSTRACT The Inflation Reduction Act (IRA) represents the largest US federal response to climate change to date. We highlight the key climate provisions and assess the act's potential economic impacts. Substantially higher investments in clean energy and electric vehicles imply that fiscal costs may be larger than projected. However, even at the high end, IRA provisions remain cost-effective. The IRA has large impacts on power sector investments and electricity prices, lowering retail electricity rates and resulting in negative prices in some wholesale markets. We find small quantitative macroeconomic effects, including a small decline in headline inflation, but macroeconomic conditions—particularly higher interest rates and materials costs—may have substantial negative effects on clean energy investment. We show that the subsidy approach in the IRA has expansionary supply-side effects relative to a carbon tax but, in a representative-agent dynamic model, is preferable to a carbon tax only in the presence of a strong learning-by-doing externality. We also discuss the economics of the industrial policy aspects of the act as well as the distributional impacts and the possible incidence of the different tax credits in the IRA.

Conflict of Interest Disclosure: Catherine Wolfram previously served as Deputy Assistant Secretary for Climate and Energy Economics, and Neil R. Mehrotra served as Deputy Assistant Secretary for Macroeconomics at the US Treasury Department. Wolfram also serves on the board of directors of Resources for the Future. 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. Other than the aforementioned, the authors are not currently an officer, director, or board member of any organization with a financial or political interest in this paper.

Brookings Papers on Economic Activity, Spring 2023: 77–157 © 2024 The Brookings Institution.

Introduction

President Biden described the climate provisions of the Inflation Reduction Act (IRA) as "the most aggressive action ever, ever, ever to confront the climate crisis" (White House 2022). Other observers similarly describe it as "the most ambitious funding ever for tackling climate change" and "the largest climate legislation in U.S. history."[1] Consistent with preliminary analysis, modeling suggests that the IRA puts the United States on track to reduce greenhouse gas (GHG) emissions 33 percent to 40 percent below 2005 levels in 2030 (Bistline, Blanford, and others 2023), which is 8 to 17 percentage points lower than without the IRA, closing the gap toward its nationally determined contribution under the Paris Agreement to halve economy-wide GHG emissions by 2030. The problem the IRA confronts is massive—reorienting the way the United States and global economies produce and consume energy. The IRA's incentives span the entire energy sector, from producers of raw materials to end-use consumers, and will set considerable new forces in motion.

The IRA is vast, and the economics profession will likely devote considerable attention over the next decade to analyzing the impacts of many of the individual programs embodied in this important piece of legislation. We offer a broad-stroke analysis of the law, summarizing the major climate-related provisions and noting some of the possible economic impacts. We focus on several major themes.

First, we discuss the possible fiscal implications of the act and note a wide range of uncertainty in the extent to which firms and households will take up the different tax credits. The Congressional Budget Office (CBO), using inputs from the Joint Committee on Taxation (JCT), estimates that over two-thirds of the fiscal costs of the climate-related provisions of the IRA ($271 billion) will be tax credits that target clean electricity production and investment, new and used electric vehicle purchases, and investments in clean energy and energy efficiency by individuals (table 1 in section I). The remaining fiscal costs ($121 billion of the $392 billion total), per CBO/JCT, will be direct expenditures on

1. The Nature Conservancy, "A 'New Day for Climate Action in the United States' as U.S. Congress Passes Historic Clean Energy and Climate Investments," https://www.nature.org/en-us/newsroom/us-house-passes-landmark-climate-bill/; and U.S. Green Building Council, "Largest Climate Bill in History Holds Good News for Green Building," https://www.usgbc.org/articles/largest-climate-bill-history-holds-good-news-green-building.

forestry and agriculture, energy loans and other financial investments, and other items. Most of the tax credits are uncapped and are a function of individual firm investment decisions and individual household consumption decisions.

We summarize evidence from the Electric Power Research Institute's US Regional Economy, Greenhouse Gas, and Energy (EPRI's US-REGEN) model in section II suggesting that initial estimates of the fiscal costs may be understated in several areas due to greater deployment of IRA-supported technologies such as clean electricity and electric vehicles. Central and higher-end estimates of tax credit expenditures range from $781 to $1,070 billion over the ten-year budget window; these are 2.9–4.0 times higher than the CBO/JCT score for comparable credits. When these tax credits are combined with direct expenditures, total budgetary effects of the IRA's climate provisions are approximately $900 to $1,200 billion cumulatively through 2031 (table 2 in section II). Even after accounting for these higher fiscal costs, IRA tax credits reduce carbon dioxide (CO_2) emissions at an average abatement cost of $36–87 per metric ton for the power sector—considerably less than recent estimates of the social cost of CO_2, with mean values between $120/t$CO_2$ and $400/t$CO_2$ in 2030 (Rennert and others 2022), depending on the near-term discount rate, even before accounting for avoided air pollution damages and other co-benefits. On the other hand, the IRA's fiscal costs may be considerably lower. We document that the costs for clean electricity generating plants, for which the IRA includes large subsidies, are more sensitive to interest rates than conventional fossil fuel generators. In addition, continued supply constraints, delays from permitting, and other factors may increase costs and reduce the pace of clean energy deployment, depressing take-up for IRA incentives. Our lower-end estimates of tax credit expenditures are about $240 billion in a scenario with higher interest rates and technology costs; total fiscal costs are slightly lower than with CBO/JCT estimates but with more limited economy-wide CO_2 reductions.

Second, we highlight potential market impacts of IRA incentives, including negative prices in wholesale electricity markets (section III). Electricity generation technologies that collect production-based tax credits will have strong incentives to operate even when wholesale prices are low and even negative (which are more common during hours when output from these resources is highest) to receive these credits. For example, a wind project may be willing to pay –$33 per megawatt-hour (MWh) (suppliers make payments when prices are negative), because it could receive as much

as $33 per MWh in tax credits. Some areas of the country are already seeing negative prices, but their prevalence will likely increase with the IRA, which can alter economic signals for market entry and exit of generators, shift incentives for locational decisions and balancing resources (e.g., energy storage, transmission), and change the economics of end-use electrification and new loads (e.g., hydrogen production, cryptocurrency mining).

Third, we discuss distributional impacts and the possible incidence of the different subsidies in the IRA. We note the extent to which the IRA may drive down retail prices for energy due to subsidies for electricity generation and investment, reflecting transfers from the federal government to consumers and clean electricity providers. In addition to potentially decreasing retail electricity prices, the IRA could lower expenditures on fossil fuels due to its incentives for end-use electrification, especially petroleum for transportation. We describe patterns in energy expenditures by income. We also present results from US-REGEN under a counterfactual scenario without IRA subsidies to inform the extent of inframarginal transfers to firms and households that would have adopted these technologies anyway.

Fourth, we consider the relationship between the IRA and the macroeconomy. To elucidate the potential macroeconomic impacts of the IRA, we develop a representative agent model of the economy, which features subsidized clean energy as an input in section IV. We show how clean energy subsidies function as a supply-side policy that boosts output, investment, wages, and labor productivity while reducing the price of electricity. These dynamic effects work to partially offset the static fiscal cost of the policy. Along the transition path, increased investment demand raises interest rates and lowers private consumption. Bottlenecks lower real clean energy investment but may raise investment expenditures and the fiscal cost of the investment tax credit as the relative price of investment in clean energy capital rises. However, the slower pace of investment under bottlenecks mitigates the rise in the real interest rate. We show how elastic labor supply and learning-by-doing externalities increase the clean energy capital stock in steady state under a subsidy policy. Even labor and domestic sourcing requirements as structured in the IRA would increase the steady-state clean energy capital stock. We also show that clean energy investment may crowd out non-energy investment in the short run but increases non-energy capital in the long run.

Fifth, we turn to a normative analysis in section V and compare the subsidies approach in the IRA to carbon pricing. Our comparison is both

conceptual and quantitative, as we derive a carbon price that would yield comparable emissions reductions over a similar time frame. Conceptually, while both policies lower the relative price of clean to fossil fuel power generation, a carbon tax raises energy prices, encouraging energy conservation but carrying negative supply-side implications for output, investment, and wages. The conservation margin means that a carbon tax results in a larger decline in emissions. In the context of our model, we define optimal policy and show that, despite its positive supply-side effects, optimal climate policy generically involves a positive carbon tax and a zero clean energy subsidy. The case for an approach centered on clean energy subsidies relies heavily on strong learning-by-doing externalities.

We describe further dimensions along which carbon taxes and subsidies differ that are not captured in our model, including fuel switching, differential carbon intensity, and impacts from usage along the intensive margin. We also compare subsidies and carbon pricing in terms of the incentives created for innovation. In this section, we also discuss the economics of some of the industrial policy aspects of the IRA, which offers higher tax credits for firms that adopt certain labor practices and buy inputs manufactured in the United States. In addition, some of the electric vehicle tax credits are only available if the vehicle meets battery sourcing requirements and North American assembly. We explain how these provisions may be addressing market failures, but if not, they may raise costs.

Last, we use outputs from the US-REGEN model in section VI as inputs into the Federal Reserve Board's US model (FRB/US) to provide quantitative evidence on the possible macroeconomic impacts of the IRA. Consistent with this, we show that new investment under the IRA, while large relative to the current level of investment in the energy sector, is comparatively small as a share of both overall investment and overall economic activity. Increases in clean power investment and household transfer income for electric vehicles and other household equipment initially increase demand before raising the capital stock and output. The movements in interest rates and unemployment are very small owing to the small size of electric power investment relative to the overall economy. Although we find that IRA investments in the baseline case are likely not large enough to meaningfully influence macroeconomic aggregates, section VII quantifies how the macroeconomic environment—including higher interest rates and rising costs of labor and materials—could have meaningful negative impacts on clean energy investment.

I. Summary of the IRA's Climate Provisions

Table 1 summarizes the major energy- and climate-related provisions in the IRA and the accompanying fiscal score (CBO 2022).[2] The fiscal score reflects estimates made by the CBO and JCT of the costs to the US government over the ten-year budget window, 2022–2031. The top panel of the table reflects tax credits, which the CBO/JCT estimate will account for about $271 billion in lost tax revenues in total through 2031. The bottom panel reflects direct expenditures, which are estimated to be $121 billion.[3] As discussed in section II, the CBO/JCT score is an initial estimate of budgetary effects of the IRA, and actual tax expenditures could be significantly larger, given how many of the tax credits are uncapped.

I.A. Tax Credits

PRODUCTION AND INVESTMENT TAX CREDITS The production and investment tax credits for clean electricity and energy storage account for about one-third of the estimated costs of the IRA's climate provisions (table 1). The production tax credit (PTC) is awarded per megawatt-hour of electricity output for the first ten years of production from qualifying low-emitting resources, while the investment tax credit (ITC) is awarded as a percentage of the investment cost.

There are two phases to each tax credit. For the first several years, until January 1, 2025, the law lists tax credit amount by type of resource (e.g., over $5/MWh for wind and solar renewable projects that do not meet labor requirements).[4] For projects that are placed into service after December 31, 2024, the law is broader and compensates any clean electricity generation

2. For more details on the climate-related provisions of the IRA, see *Building a Clean Energy Economy: A Guidebook to the Inflation Reduction Act's Investments in Clean Energy and Climate Action*, produced by the Biden administration (White House 2023a). See also Congressional Research Service (2022a).

3. Much of the reporting on the IRA has cited a total figure of $369 billion for the climate provisions, which is apparently from the initial press release on the IRA issued by the offices of Senate Democratic Leader Chuck Schumer and Senator Joe Manchin; see "Joint Statement from Leader Schumer and Senator Manchin Announcing Agreement to Add the Inflation Reduction Act of 2022 to the FY2022 Budget Reconciliation Bill and Vote in Senate Next Week," July 27, 2022, https://www.democrats.senate.gov/newsroom/press-releases/senate-majority-leader-chuck-schumer-d-ny-and-sen-joe-manchin-d-wv-on-wednesday-announced-that-they-have-struck-a-long-awaited-deal-on-legislation-that-aims-to-reform-the-tax-code-fight-climate-change-and-cut-health-care-costs.

4. All values are shown in 2022 US dollars unless otherwise noted. The base PTC level is listed in the bill text as 0.3 cents per kilowatt-hour, which is expressed in nominal terms from when the wind PTC was first applied to projects built in 1993. Many IRA tax credits include inflation adjustments over time.

Table 1. Fiscal Score of the Climate-Related Provisions of the IRA by Major Category

	Fiscal score ($ billions)
Tax credits	
Investment and production tax credits for clean electricity generation and storage	131
Production tax credit for carbon capture and sequestration	3
Nuclear power production tax credit	30
Clean fuels	19
Clean energy and efficiency incentives for individuals	37
Clean vehicles	14
Clean energy manufacturing	37
Subtotal	271
Direct expenditures	
Agricultural and forestry conservation and sequestration projects	21
Energy loans	17
Energy efficiency	11
Industrial decarbonization	5
Other (e.g., Green Bank)	66
Subtotal	121
Total	392

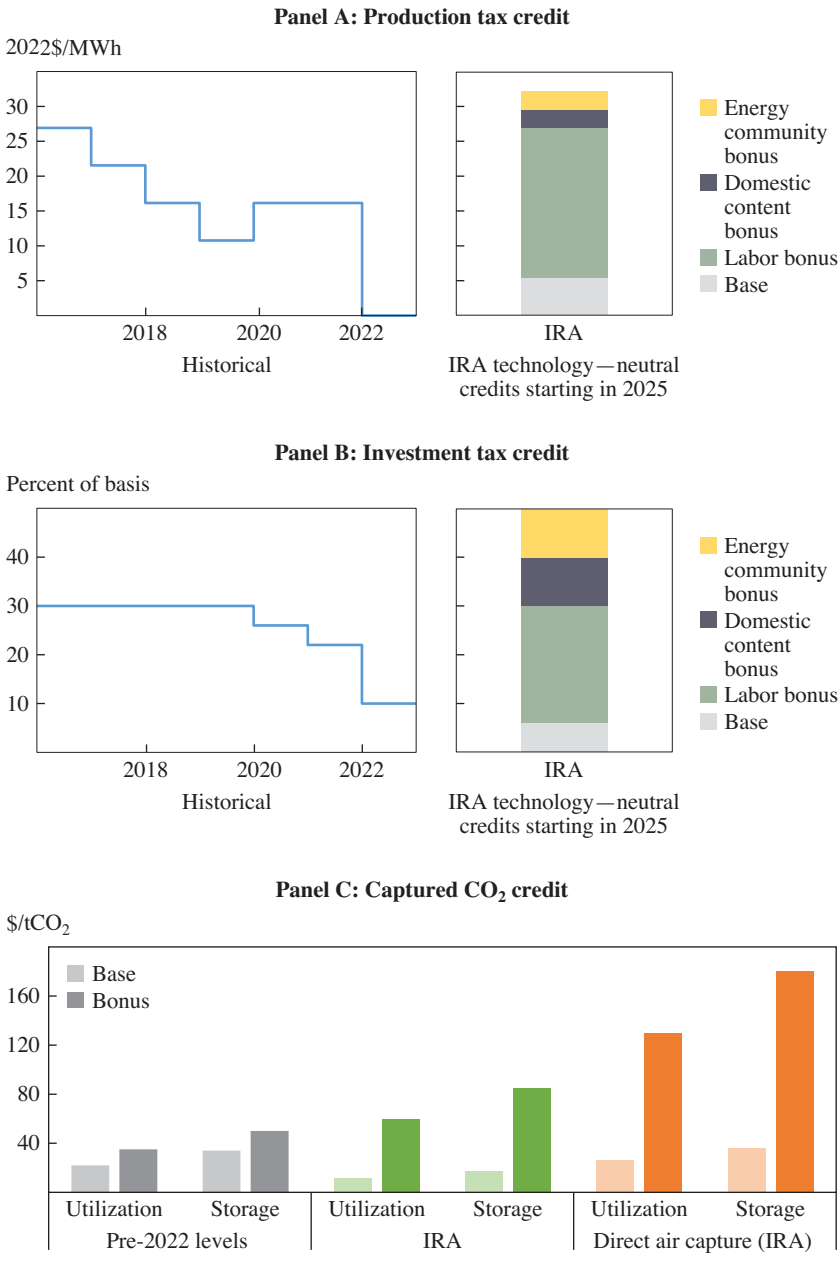
Sources: CBO (2022) and authors' calculations.

Note: Values are based on the CBO's September 7, 2022, estimates. Numbers may not sum due to rounding.

capacity, defined as one with zero GHG emissions. The tax credit is five times higher (about $27.5/MWh) for projects that meet certain labor requirements on prevailing wages and apprenticeships, as shown in figure 1. There is also a 10 percent increase in the PTC and a 10 percentage point boost to the ITC for projects that use domestically produced steel and other materials, assuming they comply with the labor requirements. Similar bonuses are available for projects that are sited in energy communities that meet specified criteria. Many of the provisions include incentives to meet similar labor, domestic content, and location-based requirements, so we describe the specifics below.

Qualifying electricity facilities are allowed to choose whether to take the PTC or the ITC, and the relative value of each credit could vary by location (e.g., due to variation in wind and solar resource quality), technology, bonus credit eligibility, and assumed capital costs. In many locations, land-based wind and solar photovoltaic (PV) have higher lifetime credit values with the PTC with the labor bonus, while offshore wind and new nuclear have higher values with the ITC (Xu and others 2022). However, if a project is eligible for both the energy community and domestic content bonuses, then the ITC could be more valuable for developers, given the higher incremental value of these bonuses under the ITC.

Figure 1. Production, Investment, and Carbon Capture Credits under the IRA Relative to Pre-IRA



Sources: Davis and others (2020); Inflation Reduction Act.

Note: The right-hand side of panels A and B reflects the technology-neutral credits starting in 2025, but the wind- and solar-specific credits available in 2023 and 2024 are identical.

PRODUCTION TAX CREDIT FOR CARBON CAPTURE AND SEQUESTRATION The IRA also expands the tax credit available since 2008 to facilities that capture carbon dioxide (section 45Q of the US Internal Revenue Code). With the IRA, facilities above a minimum size threshold that meet labor requirements are eligible for \$85/tCO_2 stored or \$60 for CO_2 utilization (figure 1, panel C). The provision applies for industrial or power-generating facilities that capture carbon dioxide from their production processes, as well as direct air capture plants built solely to capture and sequester carbon, which receive \$180 per ton of captured and stored CO_2.

Although the CBO estimates that the tax credit for carbon capture and sequestration (CCS) will cost the government only \$3.2 billion over the budget window (i.e., about 1 percent of the tax expenditures in table 1), some external modelers see substantial investment (Jenkins and others 2022; Zhao and others 2022; Bistline, Roney, and others 2023). Among those modelers that project investments in CCS, major applications include both the industrial sector and the electric power sector, though relative sequestration in each varies by model.

NUCLEAR POWER PRODUCTION TAX CREDIT The IRA adds a production tax credit through 2032 for existing nuclear power plants that meet labor and wage requirements. This credit provides up to \$15/MWh, though the magnitude of the subsidy depends on electricity revenues and whether plants already receive credits from other federal or state zero-emission credit programs. For example, the Infrastructure Investment and Jobs Act, signed into law in November 2021, created a \$6 billion program to auction grants to nuclear power plants that remain in service.

CLEAN FUELS The IRA also extends and expands credits for clean transportation and industrial fuels. As with the PTC and ITC, the legislation extends targeted tax credits for biodiesel, renewable diesel, and alternative fuels for the first several years and then replaces those with a technology neutral credit. The technology neutral credit begins in 2025 and is available through the end of 2027. The credit value is \$1/gallon if labor requirements are met, and can be increased depending on the emissions intensity of the fuel. The CBO projects the largest expenditures in this category will be a new credit for clean hydrogen (section 45V in the IRA), which can be used in transportation, industrial, and power generation applications, and the magnitude of these hydrogen subsidies depends on the emissions intensity of production.[5] The IRA also adds a tax credit for sustainable aviation fuels

5. Credits for electrolytic hydrogen may be combined with PTC and ITC incentives for clean electricity production, which is an input to hydrogen production.

of $1.75/gallon with the labor bonus, although CBO estimates, consistent with those of outside modelers, reflect relatively low take-up of this credit (Bistline, Blanford, and others 2023). Unlike the power sector PTC, the tax credits apply to all qualifying fuels produced that year, whether they are from a new facility or not.

CLEAN ENERGY AND EFFICIENCY INCENTIVES FOR INDIVIDUALS The CBO estimates that individuals will make use of almost $40 billion in tax credits for clean energy and energy efficiency investments (table 1).[6] Individual taxpayers can receive credits for their investments in equipment, including home solar; battery storage; solar water heating; small wind energy; energy efficient insulation, windows, and doors; electric heat pumps; and home energy audits and electric panel upgrades necessary for other efficiency improvements. The amount of rebate can vary based on the energy savings, building type, and household income. There are caps on the amounts an individual taxpayer can claim for specific investments (e.g., $150 for a home energy audit and $2,000 for a heat pump) and on total annual credits, but there are no caps on the total amount of credits. Unlike the more commercially oriented credits discussed thus far, these credits do not offer bonuses for using a particular category of labor.

CLEAN VEHICLES The IRA allows taxpayer credits up to $7,500 for the purchase of a new electric or hydrogen fuel cell vehicle if several conditions are met. The conditions include the following: the final assembly of the vehicle must take place in North America; a share of both the critical minerals and the battery components must come from North America (or, in the case of critical minerals, a country with which the United States has a free-trade agreement), and the share escalates over time after 2024; and both the vehicle manufacturer's suggested retail price (MSRP) and the taxpayer's income must be below specified limits. Of the total credit, $3,750 is tied to meeting the battery components requirement and $3,750 to the critical minerals requirement. Treasury guidance issued in March 2023 confirmed that companies leasing vehicles to consumers may claim the commercial clean vehicle credits, which can provide $7,500 without stringent requirements on battery sourcing or caps on MSRP or income eligibility (US Department of the Treasury 2023a). More broadly, the commercial clean vehicle tax credits provide up to $7,500 for vehicles less than 14,000 pounds and $40,000 for larger vehicles (or 30 percent of the purchase price or incremental cost of an internal combustion engine replacement, whichever is lower).

6. Less than $0.5 billion of this line item is for energy efficiency investments in commercial buildings.

The IRA also introduces a $4,000 credit (or 30 percent of the vehicle price, whichever is smaller) for the purchase of a previously owned electric vehicle as long as the vehicle is more than two years old, the buyer meets income requirements, and the selling price is below $25,000. Both the income and selling price thresholds are considerably lower than those for new electric vehicles. Finally, the IRA extends tax credits for individual taxpayers, who can, regardless of income, deduct up to 30 percent or $1,000 for home charging. Businesses can also claim tax credits for electric vehicle charger installations of 30 percent or (as long as they satisfy labor requirements) up to $100,000.

CLEAN ENERGY MANUFACTURING The IRA extends and expands tax credits for retrofits or new construction of certain energy manufacturing facilities, such as facilities that produce energy storage systems or electrolyzers. It features a 30 percent credit capped at $10 billion that applies to a range of clean energy technologies and an uncapped credit per unit of production for several specific wind, battery, and solar components (for example, $12 per square meter of photovoltaic wafer or $3 per kilogram of solar-grade polysilicon). The CBO estimates that the bulk of the tax expenditures will be through the uncapped provision, though analysis by Credit Suisse indicates that manufacturing credits could be many times the CBO/JCT estimate (Jiang and others 2022).

I.B. Common Features of the Tax Credits

Most of the IRA tax credits are not capped, so the CBO/JCT estimates summarized in table 1 are subject to considerable uncertainty. In section II, we present examples of models that have come up with widely varying estimates of the fiscal costs of the credits. In some cases, such as production-based credits for low-carbon electricity, carbon capture, and clean hydrogen, lower fiscal costs would be driven by lower than anticipated deployment, meaning that emissions would not come down as quickly as hoped. For the ITC, however, lower budgetary effects could also be driven by lower than expected investment costs—for example, if the costs of renewable or storage technologies fell faster than expected. Budgetary effects of the tax credits also could be lower if bonus credit uptake is limited.

The IRA does two things to make the tax credits easier to use: it makes some of them direct pay and some of them transferable.[7] Direct pay

7. Some credits are transferable but not eligible for direct pay, such as the personal electric vehicle credit. Some are eligible for direct pay but not transferable, such as the credit for commercial clean vehicles. Many are both, including the PTC and ITC for renewable electricity. Credits for alternative fuels are neither eligible for direct pay nor transferable.

essentially transforms the tax credit into a grant and means that entities such as nonprofits and state and local governments are eligible to receive them. If a tax credit is transferable, a taxpayer can transfer the credit to an unrelated party in exchange for cash. This means that if a provider (e.g., a solar power project developer) has a tax bill that is too small to absorb the credits, it can transfer the credits to a taxpayer who can use them, that is, one whose tax bill is larger than the value of the credits. This transferability provision did not exist before the IRA; so, for example, renewable developers had to form partnerships with taxpayers who had the ability to absorb the credits (so-called tax equity investors). As another example, the IRA specifies that beginning in 2024, the taxpayers may elect to transfer the electric vehicle credit to the dealer, meaning that the credit works like a point-of-sale rebate.

Almost all the credits include substantial bonuses for projects that use domestic content, are located in low-income or energy communities, and meet certain labor requirements (figure 1). Companies can comply with the labor requirements if they pay prevailing wages during construction and repair and if qualified registered apprentices provide more than a threshold share of labor hours, where the threshold increases over time. This bonus is lucrative for projects eligible for the PTC and ITC, increasing values by five times relative to the base rate. The domestic content provisions typically increase over time, presumably to allow US manufacturers the opportunity to scale production capacity.

The energy communities bonus provides an additional 10 percent for the PTC or 10 percentage points for the ITC if any of three criteria are met related to brownfield sites, employment and tax revenue from fossil fuels, or coal mine or plant closures.[8] Eligibility for energy community bonuses could affect bonus uptake, siting decisions for projects, and the ability to direct IRA funds toward areas experiencing the most acute impacts from lower fossil fuel consumption and production. The broad geographical coverage of statistical areas and census tracts under energy communities definitions likely means that large land areas may be eligible for these bonuses,

8. The IRA specifies a "0.17 percent or greater direct employment or 25 percent or greater local tax revenues related to the extraction, processing, transport, or storage of coal, oil, or natural gas" and an "unemployment rate at or above the national average unemployment rate for the previous year" (Inflation Reduction Act of 2022, H.R. 5376, 117th Cong. (2021–2022), Pub. L. 117–169 (August 16, 2022), https://www.congress.gov/bill/117th-congress/house-bill/5376/).

covering 42–50 percent of US land area according to initial estimates (Raimi and Pesek 2022), which is consistent with Treasury and Internal Revenue Service (IRS) guidance. Even areas that are geographically distant from communities that are dependent on fossil fuels for employment and government revenue may be eligible (US Department of the Treasury 2023b). This coarse geographical targeting means that IRA benefits also go to regions with limited fossil fuel dependence, while at the same time, the unemployment criterion may exclude areas with some of the highest dependence on fossil fuels, including Colorado, Louisiana, North Dakota, West Texas, and Wyoming (Raimi and Pesek 2022). Additionally, the binary eligibility rule for the energy communities bonus ignores the heterogeneity of fossil fuel dependence, which lowers the cost-effectiveness of these provisions in achieving "just transition" objectives. The modeling in section II illustrates that the eligibility for the energy communities bonus can increase the deployment of wind, solar, and other IRA-qualified resources, though the geographical allocation of this capacity does not necessarily align with communities dependent on fossil fuel.

The credits are available over different time periods. The production and investment tax credits begin to phase down either in 2032 or at the point when power sector emissions reach 25 percent of their 2022 emissions, whichever is later. This emissions-based eligibility threshold is a novel feature of the IRA and could imply that qualifying clean electricity resources may receive tax credits well into the 2030s and potentially longer, which has associated budgetary impacts. Production-based tax credits for hydrogen and electricity apply to projects placed in service through at least 2032 and continue for ten years after the project begins claiming the credit. Credits for CO_2 capture continue for twelve years after the project begins claiming the credit. The electric vehicle tax credits are generally available through 2032, which is the end of the ten-year budget reconciliation period.

I.C. Direct Expenditures

AGRICULTURAL AND FORESTRY CONSERVATION AND SEQUESTRATION PROJECTS

The IRA provides more than $20 billion for agricultural and forestry conservation programs. Much of the agricultural funding flows through existing conservation programs, though the IRA expands funding for them significantly. For instance, $8.45 billion is directed to the Environmental Quality Incentives Program (EQIP) for practices that improve carbon storage in soil or decrease GHG emissions. Funding for forestry programs

would target hazardous fuel reduction projects, vegetation management projects, inventories of old-growth forests, and other measures.[9]

ENERGY LOANS The IRA increases existing loan program authority of the US Department of Energy's Loan Programs Office by about $100 billion and creates a new Energy Infrastructure Reinvestment Program, which aims to accelerate retooling and replacing emissions-intensive energy infrastructure. The IRA also increases funding for several existing programs that aim to encourage farmers and rural landowners to purchase renewable energy systems. It also provides almost $10 billion to encourage rural electric cooperatives to invest in renewables and other low-carbon energy projects.

ENERGY EFFICIENCY The IRA includes over $10 billion in direct expenditures for energy efficiency programs, including a new Department of Energy program to award grants to state energy offices to develop whole-house energy-saving retrofit programs.[10] It also increases funding for energy efficiency under an existing Department of Housing and Urban Development affordable housing program.

INDUSTRIAL DECARBONIZATION The law funds a new program at the Department of Energy to support industrial facilities in emissions-intensive industries that complete demonstration and deployment projects to reduce emissions. Although projections before the IRA suggest that the industrial sector will be responsible for over one-quarter of emissions by 2030, the IRA allocates only $5 billion for emissions reductions in the sector. Some models suggest that clean hydrogen and CCS will be useful in the industrial sector, so the total subsidies for the sector may be larger than $5 billion, but still considerably smaller than subsidies for the electric power and transportation sectors (Bistline, Blanford, and others 2023).

OTHER Notable expenditures in the "Other" category in table 1 include $27 billion for the Environmental Protection Agency (EPA) to run the Greenhouse Gas Reduction Fund, which will award competitive grants with an emphasis on clean energy projects that benefit low-income and disadvantaged communities. (This is sometimes described as the US government's "green bank," as much of the funding will support nonprofit organizations that provide financial or technical assistance to local clean energy projects.)

9. See Mahajan and others (2022) for analysis of these provisions. They project that these provisions lead to CO_2 emissions from land use change and forestry in 2030 of -744 $MtCO_2$ equivalent/year (i.e., net removals) instead of -707 $MtCO_2$ equivalent/year in the counterfactual reference without the IRA. See Congressional Research Service (2022a) for more details on the agricultural and forestry programs.

10. US Department of Energy, "About the Home Energy Rebates," https://www.energy.gov/scep/home-energy-rebate-programs.

The IRA includes a Methane Emissions Reduction Program to establish a charge on methane emissions from specified sources, beginning at $900 per metric ton of methane and increasing to $1,500 after two years. This fee equates to about $36 and $60 per $MtCO_2$ equivalent, although the precise number depends on the assumed CO_2 equivalence of methane (Congressional Research Service 2022b). The fee will not be assessed if states adopt EPA regulations.

I.D. Comparing the IRA to Pre-2022 Programs

In many cases, the IRA extended and expanded existing programs. In some cases, the IRA introduced entirely new programs.

As an example of enhancements to existing tax credits, figure 1 depicts the production and investment tax credits that existed for wind, solar, and CCS before the IRA. Panel A summarizes the credits available to wind generators through the production tax credit, which had been as high as $27/MWh in 2016 (2022$) but expired on December 31, 2021. Before the IRA, solar projects were eligible for the ITC but not the PTC. The bars on the right break down the IRA credits, showing the large bonus credits available for plants that meet the labor requirements. Panel B of figure 1 compares historical ITC values with IRA ones. Finally, panel C summarizes credits available for captured CO_2. Before the IRA, industrial and power-generating facilities were eligible for $50/$tCO_2$ for CO_2 storage and $35/$tCO_2$ for utilization. The updated IRA credits for CO_2 capture also include higher credit levels for direct air capture, which can be up to $180/$tCO_2$ for projects with the labor bonus.

The longevity of the tax credits under the IRA also provides renewable project developers with more certainty than they had in the past. For example, the Union of Concerned Scientists noted in 2015 that, since its inception with the Energy Policy Act of 1992, the production tax credit expired six times. Though it was subsequently extended by Congress each time, they conclude that "this 'on-again/off-again' status has resulted in a boom-bust cycle of development. In the years following expiration, installations dropped between 76 and 93 percent."[11] Timing projects to ensure that they qualify for tax credits may increase costs and risk for renewables developers.

Examples of expansions under the IRA include extending the investment tax credit to stand-alone energy storage projects, which previously were only eligible if they were colocated with solar power facilities.

11. Union of Concerned Scientists, "Production Tax Credit for Renewable Energy," July 15, 2008, updated February 9, 2015, https://www.ucsusa.org/resources/production-tax-credit-renewable-energy.

In some cases, the IRA introduced brand new programs, including the EPA's Greenhouse Gas Reduction Fund, tax credits for advanced manufacturing of clean energy inputs, and tax credits for commercial vehicles. And many of the labor market, domestic content, and energy community credits are new.

II. Fiscal Implications

II.A. Initial CBO/JCT Score

Initial estimates by the CBO and JCT (CBO 2022) indicate that the entire climate and non-climate provisions of the IRA would increase federal tax revenue by $58 billion on net over the ten-year budget window (i.e., through 2031).[12] Increases in fiscal costs from tax credits and direct spending would be more than offset by increases in revenues from alternative minimum taxes on large corporations, excise taxes on stock buybacks, and increased enforcement of extant taxes. According to this CBO/JCT estimate, costs to the US government for energy- and climate-related tax credits are $271 billion and $121 billion for direct expenditures through 2031 (table 1).

As discussed in section I, large shares of spending are allocated for IRA provisions where tax credits are uncapped, which are investment- or production-based. This means that actual federal tax expenditures and budgetary effects might be significantly more than initially estimated.

II.B. US-REGEN Score

Here, we use the energy-economic model US-REGEN to estimate the budgetary effects of core energy-related IRA provisions.[13] US-REGEN

12. For a summary of CBO/JCT estimates, see CBO (2022). For an overview of the CBO's general approach to preparing cost estimates, see CBO (2018); for the JCT's approach, see JCT (2019). These quick turnaround analyses of proposed legislation estimate changes to federal spending and revenues for most bills approved by committee in the House or Senate. For the IRA's power and transport provisions, JCT revenue estimates come from baseline projections, often based on US Energy Information Administration projections, combined with tax data and assumed elasticities.

13. EPRI's US Regional Economy, Greenhouse Gas, and Energy (US-REGEN) model features regional disaggregation and technological detail of the power sector and linkages to other economic sectors. Recent peer-reviewed articles and reports can be found at EPRI, "Models," https://esca.epri.com/models.html, and detailed model documentation at "US-REGEN Documentation," https://us-regen-docs.epri.com/. For descriptions of IRA implementation, see Bistline, Roney, and others (2023). The analysis was conducted in February 2023 and does not consider changes in policy or guidance after that time.

brings technological, temporal, spatial, and cross-sector detail, which can influence the economics of supply- and demand-side resources in the energy system and consequently the budgetary effects of IRA provisions. The electric sector model is an intertemporal optimization of capacity planning and dispatch with simultaneous investments in generation, energy storage, transmission, hydrogen production, and carbon removal.[14] Demand for electric vehicles, heat pumps, and other building equipment comes from separate logit choice models that translate relative costs of ownership across technologies into equilibrium market shares.[15]

This analysis suggests that government expenditures under the IRA may be significantly larger than initial estimates based on higher tax credit uptake and deployment of clean electricity, carbon capture, and electric vehicles. Total fiscal costs of tax credits in US-REGEN are estimated to be over $780 billion by 2031—nearly three times the CBO value for comparable credits. Figure 2 compares total fiscal costs (i.e., cumulative expenditures) of IRA tax credits in the US-REGEN analysis with the CBO/JCT score for select IRA provisions over time. When these tax credits are combined with direct expenditures, total budgetary effects of the IRA's climate provisions total $901 billion through 2031.

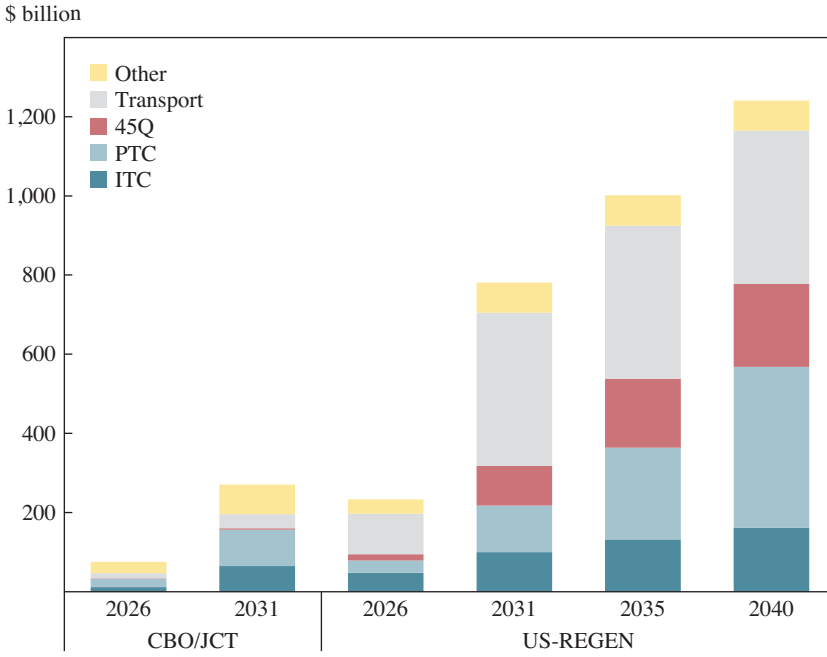
Tax credits in the electric sector are projected to be $320 billion cumulatively through 2031 by US-REGEN. One of the largest differences between CBO/JCT and US-REGEN estimates is 45Q credits for captured CO_2. The CBO/JCT score indicates a total of $3.2 billion in 45Q credits through 2031 (and just $0.3 billion annually in 2031). In contrast, credits for captured CO_2 in US-REGEN total $100 billion through 2031.[16] US-REGEN estimates could be conservative, since they reflect only credits used in the power sector. BloombergNEF estimates cumulative 45Q credits could be

14. US-REGEN uses a unique approach for selecting intra-annual segments to more accurately characterize the economics of variable renewables, energy storage, and dispatchable capacity (Blanford and others 2018). This algorithm allows the model to capture features such as diminishing marginal returns for higher wind and solar deployment, the value of firms' low-emitting technologies under deep decarbonization, and chronological system operations for short- and long-duration energy storage options.

15. The model includes segmentation of firms and consumers to capture differences that are relevant to technology choice, including household location, vehicle ownership, driving intensity, building type, and charging access for passenger vehicle decisions. Hourly electricity profiles for vehicle charging are based on exogenous charging patterns that vary by driver type, day type, temperature, and location and include flexibility through deferrable charging (Bistline, Roney, and others 2021). US-REGEN assumes perfectly competitive markets, meaning that tax credits and taxes are passed through to consumers.

16. Note that US-REGEN is typically run in five-year time steps. However, for comparability with the CBO/JCT score, we interpolate results to show 2026 and 2031 values.

Figure 2. Estimates of Cumulative (Undiscounted) Fiscal Costs from IRA Tax Credits by Provision



Sources: CBO (2022) and authors' calculations.

Note: CBO/JCT scores are based on September 7, 2022, estimates and look at budgetary effects from 2022 to 2031. Values are shown in nominal terms. The "Other" category includes additional end-use incentives (e.g., credits for heat pumps) and manufacturing. The 45Q credits are for captured CO_2.

over $100 billion, as carbon capture and sequestration (CCS) could become economical for several point-source CO_2 applications outside of the power sector, including natural gas processing, ethanol, ammonia, and cement (Attwood 2022).

Note that the CBO/JCT score examines effects only across a ten-year budget window, and most IRA incentives are available for ten years from their construction date (i.e., a project constructed in the early 2030s may receive credits into the 2040s). US-REGEN estimates that the electric sector tax credits will sum to $780 billion through 2040, 63 percent of the total (figure 2). Cumulative credits for captured CO_2 are estimated to be $210 billion by 2040. For a hypothetical budget window from 2031 to 2040, electric sector tax credits would be $460 billion, nearly three times the comparable CBO/JCT values for the initial ten-year period (about $160 billion).

Fifty-nine percent of aggregate spending through 2031 in US-REGEN comes from demand-side credits for electric vehicles, heat pumps, and energy efficiency upgrades. In particular, subsidies to promote electric vehicle sales dominate end-use credits and are significantly higher than initial estimates. Total fiscal costs of clean vehicle credits in US-REGEN are $390 billion through 2031, which is more than an order of magnitude greater than CBO estimates of $14 billion. If the US Energy Information Administration's (EIA) estimates are used for electric vehicle deployment, the CBO/JCT score suggests that the average eligible share of vehicles for the full credit is 12 percent.[17] On the other hand, if all new electric vehicles in 2030 are assumed to be eligible for the full credit, the CBO/JCT score implies an electric vehicle sales share of 1 percent.

Because electric vehicles can have lower total costs of ownership for many households before tax credits (Burnham and others 2021), 73 percent of electric vehicles sold in 2030 would have occurred in the counterfactual without IRA incentives in US-REGEN, reducing the efficiency of these tax credits.[18] Figure 3 indicates that the IRA increases the electric vehicle share of new vehicle sales by 12 percentage points in 2030, from 32 percent to 44 percent. There is considerable uncertainty about the fiscal cost of consumer tax credits for electric vehicles since, as section I.A describes, magnitudes of these credits are tied to battery components and critical minerals requirements, various eligibility restrictions, and domestic manufacturing incentives. These factors depend on Treasury guidance, firms' responses to passenger clean vehicle credits and manufacturing tax credits, and consumer purchasing decisions.[19]

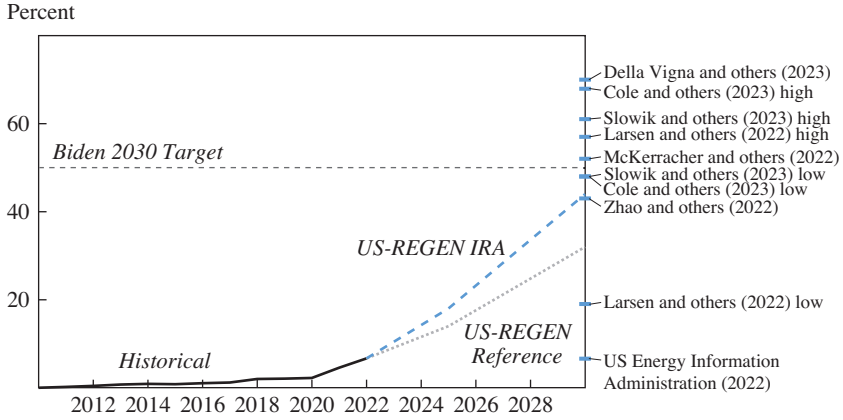
Tax expenditures increase over time, especially in the electric sector. Figure 2 illustrates how fiscal costs of tax credits in 2031 total $781 billion but reach over $1.2 trillion by 2040. End-use incentives are larger in 2031,

17. The US Energy Information Administration (2022) projects that electric vehicle sales shares are 6.6 percent in 2030, declining from a 6.7 percent sales share in 2022 (see figure 3).

18. Electric vehicle shares account for range anxiety, household heterogeneity, preferences for internal combustion engine vehicles, and other factors in US-REGEN's logit choice model of passenger vehicles, as described in the model documentation at EPRI, "Transportation," https://us-regen-docs.epri.com/v2021a/assumptions/transportation.html.

19. For US-REGEN scenarios, new light-duty vehicle sales are assumed to have an increasing average incentive value for clean vehicle credits (section 30D in the Internal Revenue code) and advanced manufacturing production credits (section 45X) over time. Assumptions about battery assembly, critical materials sourcing, income and sales price eligibility, final vehicle assembly, and battery manufacturing credits lead to an average incentive value for new electric vehicles of $3,750 in 2025 and $7,500 in 2030, which are similar to the moderate scenario in Slowik and others (2023).

Figure 3. Electric Vehicle Share of New Passenger Vehicle Sales



Sources: International Energy Agency (2022; historical data) and authors' calculations.

Note: Electric vehicles include battery and plug-in hybrid vehicles. US-REGEN scenarios with and without the IRA are compared with recent estimates of electrification shares of new vehicle sales under the IRA.

but since these credits generally expire around 2031, power sector credits lead cumulative spending by 2040. The power sector PTC and ITC could last longer, until electricity emissions reach 25 percent of 2022 levels, which could mean that credits could remain in place for over two decades and that fiscal implications of IRA credits are largest after the initial ten-year window.[20] Projections for tax expenditures over time can vary based on uptake of ITC vis-à-vis PTC (i.e., where the former are front-loaded and the latter are payouts over time), bonus eligibility, and timing of investments.

II.C. Comparisons with Other Estimates of the IRA's Fiscal Costs

This finding that fiscal costs under the IRA are uncertain and may be much larger than initial CBO/JCT estimates is reflected in other studies.

An analysis by Credit Suisse points to greater climate spending in several areas, especially for advanced manufacturing credits. The authors of the report project tax expenditures of $250 billion for these credits supporting solar, wind, and battery supply chains, which is eight times higher than CBO estimates (Jiang and others 2022).

20. This threshold is approximately 380 $MtCO_2$ equivalent/year, based on preliminary estimates of 2022 emissions by the Rhodium Group (Rivera and others 2023). IRA scenarios in US-REGEN generally do not reach emissions levels below this threshold until after 2040.

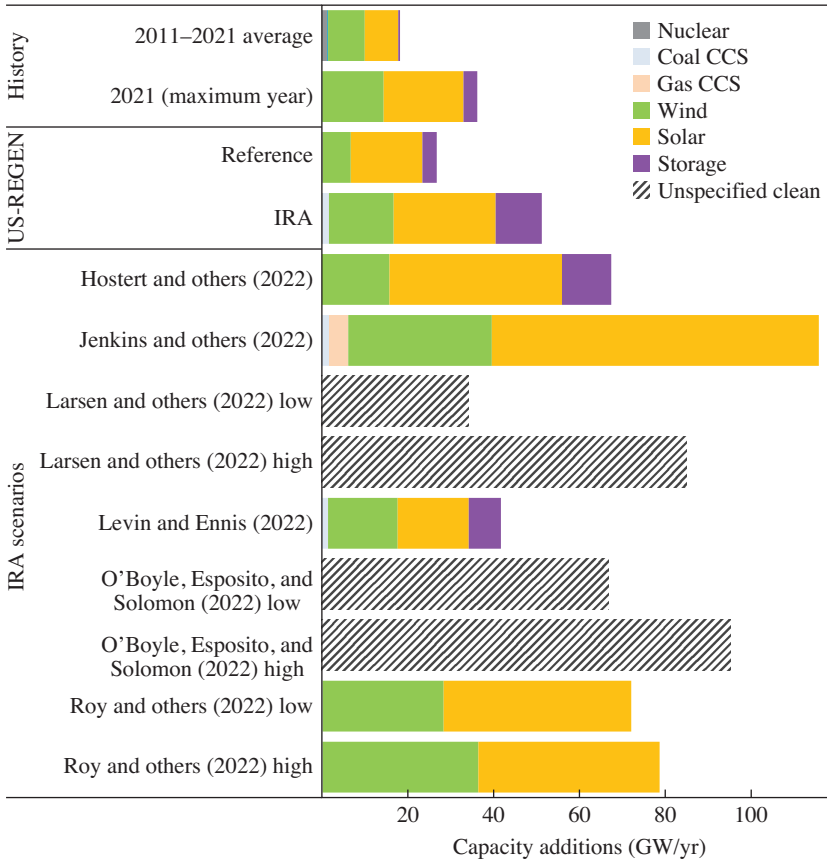
An analysis by Goldman Sachs (Della Vigna and others 2023) estimates that government spending could be nearly $1.2 trillion over the next decade, including $393 billion for transport electrification and $274 billion for clean electricity.

Cole and others (2023) estimate cumulative federal tax expenditures for light-duty vehicles to be $451 billion through 2031, comparable to the $390 billion in US-REGEN, which is more than an order of magnitude greater than the $14 billion CBO value. They also estimate that 40–57 percent of this spending would be inframarginal transfers to consumers who would have purchased in the counterfactual without IRA incentives. Figure 3 compares the estimates by Cole and others (2023) for new sales of electric vehicles under the IRA with US-REGEN values and several other estimates from the literature (Slowik and others 2023; Zhao and others 2022; McKerracher and others 2022; Larsen and others 2022). These shares span a wide range—from 19 to 70 percent—though many are above the US-REGEN value of 44 percent, suggesting that the budgetary effects in figure 2 could be conservative.

Power sector capacity additions vary based on IRA implementation and scenario assumptions (e.g., projections of capital costs, fuel prices, and other policies and incentives). Figure 4 compares US-REGEN additions of low-emitting capacity (including renewables, nuclear, CCS-equipped fossil fuels, and energy storage) with other public estimates of the IRA's impacts (Roy and others 2022; Levin and Ennis 2022; Hostert and others 2022; O'Boyle, Esposito, and Solomon 2022; Jenkins and others 2022; Larsen and others 2022). Average annual capacity additions under the IRA range from 34 GW/year to nearly 120 GW/year, which suggests that the US-REGEN estimates in earlier sections could underestimate fiscal costs of IRA provisions.

The comparison of electric sector additions in figure 4 also illustrates how although IRA tax credits accelerate clean electricity deployment, there are still considerable additions of solar, wind, and energy storage in the counterfactual without the IRA. Technological change has led to rapid cost declines for solar power, wind power, and battery storage over the past decade (see figure 18), and the expectation of future declines in costs of these resources (see figure 8) implies that a portion of the power sector tax credits will be inframarginal transfers to firms that would have adopted these technologies even without the IRA (similar to clean vehicle credits), which aligns with qualitative insights from earlier tax credit analysis (Stock and Stuart 2021). Average annual additions of low-emitting capacity in US-REGEN is 51 GW/year with the IRA and 27 GW/year in the counterfactual reference,

Figure 4. Average Annual Capacity Additions by Low-Emitting Generation and Energy Storage Technologies



Sources: US Energy Information Administration and authors' calculations.

Note: Historical values come from US Energy Information Administration, "Form EIA-860 Detailed Data with Previous Form Data (EIA-860A/860B)," https://www.eia.gov/electricity/data/eia860/, and show average additions over the past decade and the year of maximum deployment to date (2021). US-REGEN model outputs show build rates with and without the IRA through 2035. These scenarios are compared with estimates in the literature for deployment under the IRA through 2035, where available. Note that several studies do not report energy storage capacity or technology-specific capacity additions (aggregate values are shown in stripes).

indicating that over half of these additions would have occurred without IRA tax credits. Nonetheless, US-REGEN modeling suggests that the IRA incentives are highly cost-effective, as we discuss below.

Figure 4 also illustrates the broad range in possible electric sector additions under the IRA, even across models with similar IRA implementations (Bistline, Blanford, and others 2023). These cross-model differences in power sector outcomes are tied to model structure, input assumptions, and IRA representations. In particular, temporal resolution (i.e., the number of intra-annual periods represented for investment and dispatch decisions) and assumed discount rates (section III.A) alter the economics of low-emitting resources (Bistline 2021; Lonergan and others 2023).[21]

II.D. IRA Emissions Impacts and Implied Abatement Costs

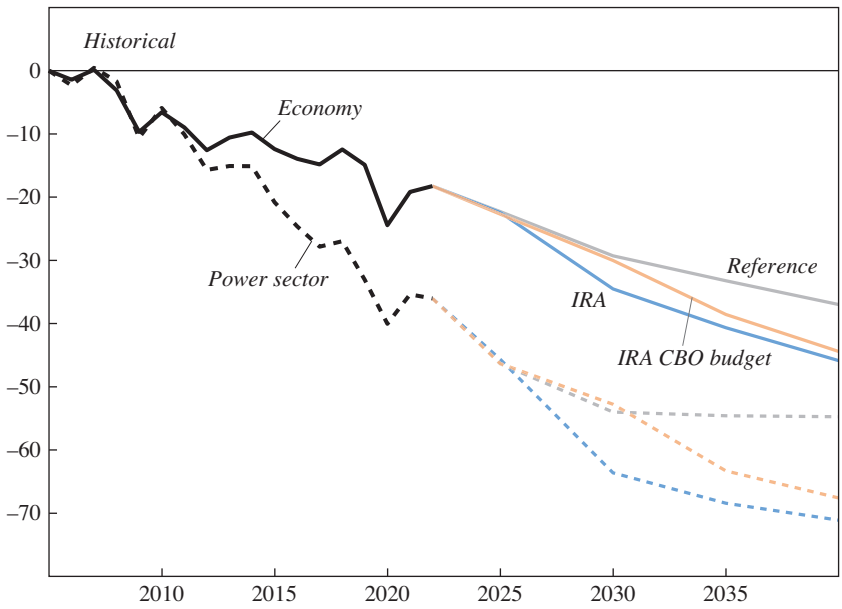
Although the CBO/JCT score does not provide carbon reduction estimates, the initial announcement of the IRA indicated that it would "reduce carbon emissions by roughly 40 percent by 2030" (Senate Democrats 2022). Figure 5 shows CO_2 emissions over time for the US-REGEN scenarios, indicating economy-wide reductions with the IRA of 35 percent by 2030 (from 2005 levels) and 41 percent by 2035 (compared to 29 percent and 33 percent, respectively, in the reference scenario without the IRA).[22] Large shares of IRA-induced emissions reductions beyond reference levels come from the electric sector, which decreases its emissions 64 percent by 2030 (compared to 54 percent in the reference). If a 40 percent reduction in economy-wide emissions were reached by 2030, fiscal costs of the tax credits would exceed the US-REGEN estimates in figure 2, making them even higher than initial CBO estimates (the sensitivity in the next subsection indicates that tax credit expenditures would be nearly $1,100 billion through 2031 if a 40 percent reduction is reached).

We also conduct a sensitivity in US-REGEN to evaluate emissions reductions if fiscal costs over the ten-year budget window are constrained to the CBO/JCT score values for power sector tax credits (table 1). Figure 5

21. For instance, US-REGEN has one of the highest temporal resolutions of the models represented in figure 4 but lower deployment, whereas one of the models with highest deployment has a single annual time slice (Energy Innovation 2021).

22. This 6 percentage point reduction in energy CO_2 emissions is comparable to the US Energy Information Administration's reduction between its reference case with the IRA and its no-IRA scenario, which is 7 percentage points—from 26 percent below 2005 by 2030 to 33 percent with the IRA (US Energy Information Administration 2023).

Figure 5. Economy-Wide and Electric Sector CO_2 Emissions over Time



Sources: Environmental Protection Agency; CBO; and authors' calculations.

Note: Values are based on US-REGEN modeled scenarios with IRA incentives, a counterfactual reference without the IRA, and an IRA scenario with a constraint that fiscal costs match CBO values through 2030. Historical values come from the US Environmental Protection Agency, "Inventory of U.S. Greenhouse Gas Emissions and Sinks," https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks.

shows how economy-wide emissions reductions reach 30 percent below 2005 only by 2030 in that case.[23]

We calculate emissions reductions in the US economy and do not model impacts on the rest of the world. All of the models summarized in figure 4 take a similar approach. This implicitly assumes that reductions in the United States will not lead to increased emissions outside the United States (i.e., that there will not be meaningful emissions "leakage"). The existing empirical estimates and model-based studies suggest that emissions leakage is limited (Grubb and others 2022), and these studies are based on climate mitigation approaches that impose costs on domestic industry, such

23. Power sector emissions in a scenario with IRA incentives that are capped at the CBO values exceed reference levels, since renewables and CCS deployment are lower than in the IRA scenario but overall generation is higher than the reference case due to the additional end-use electrification from IRA incentives.

as carbon pricing. Since the IRA will subsidize clean production, emissions leakage is likely to be even more limited. We discuss possible technological and policy spillovers to the rest of the world below.

Our estimated emissions reductions can be used to calculate implied abatement costs associated with IRA tax credits. In particular, the cost per metric ton of CO_2 reduced from IRA tax credits in the power sector is based on changes in total system costs over time (including all private sector and government costs that would not have been borne without the IRA) divided by emissions impacts from the IRA (which is the difference between the IRA and reference scenarios in figure 5).[24] Cumulative incremental costs under the IRA are $420–570 billion across the model's time horizon, depending on the discount rate—ranging from 1.5 percent to 3 percent—with cumulative reductions of 9.4 billion metric tons of CO_2. As shown in figure 6, IRA tax credits reduce CO_2 emissions at an average abatement cost of $45–61 per metric ton for the power sector—considerably less than recent estimates of the social cost of CO_2—with mean values between $120–400/t$CO_2$ in 2030 (Rennert and others 2022), depending on the near-term discount rate, even before accounting for avoided air pollution damages and other co-benefits.[25] The finding that the IRA's average abatement costs are likely below updated social cost of carbon estimates holds even with the higher abatement costs used in section II.E and inframarginal transfers noted in section II.C.

II.E. Lower and Higher Fiscal Cost Sensitivities

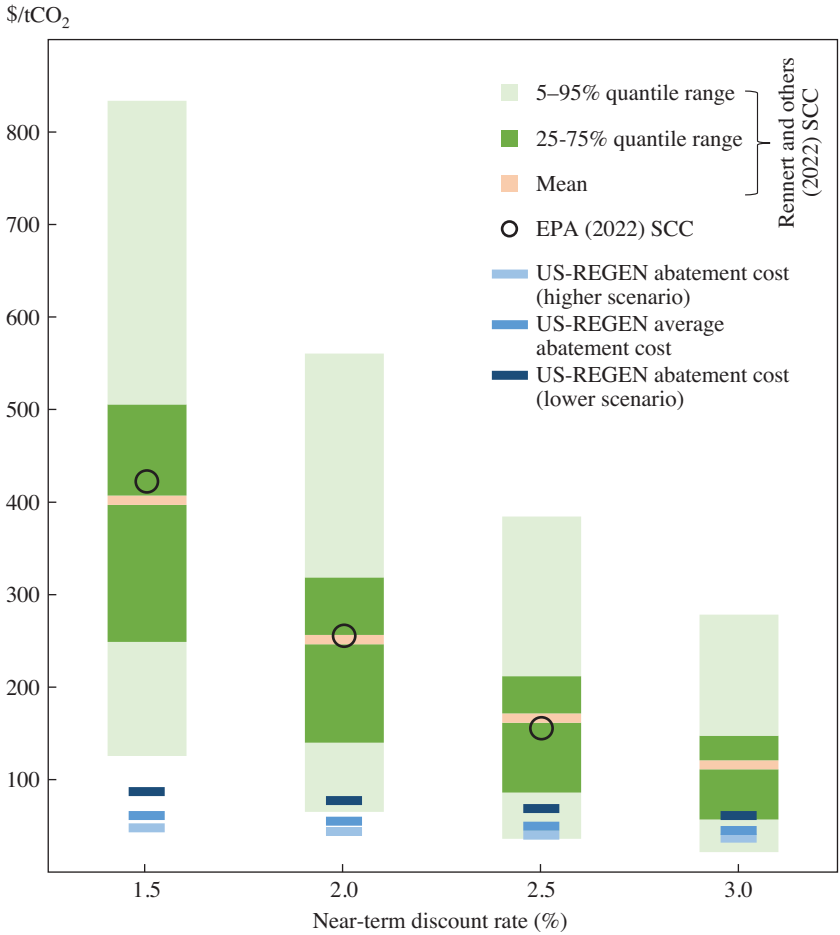
Given uncertainties in the planning environment, we include two sensitivities to investigate the potential range of the IRA's fiscal costs, in addition to the central case presented earlier.[26] A *lower fiscal costs* scenario

24. Similar to the definitions used by Stock and Stuart (2021) and Greenstone and others (2022), this definition includes incremental expenditures on capital costs (including generation, energy storage, transmission, and distribution), fuel costs, and maintenance costs, and includes both public and private expenditures. Abatement costs include changes in cost and emissions through the model horizon of 2050, since incentives can shift cash flows far into the future, given the long-lived nature of many assets and the potential for capital-intensive, low-emitting resources to lower operational costs over time. Comparisons use discounted costs and undiscounted emissions, due in part to the comparability with emissions-equivalent carbon price in section V.E, where the shadow price on the annual emissions cap constraint is implicitly using discounted costs (since the objective function is the net present value of system costs) and undiscounted emissions. If both costs and emissions are discounted, average abatement costs are $88–94/t$CO_2$.

25. Note that average abatement costs would be higher if distortions associated with financing the incentives were included (Finkelstein and Hendren 2020).

26. These illustrative scenarios should not be interpreted as bookends of possible scenarios or as predictions of likely outcomes.

Figure 6. Comparison of Social Cost of CO_2 Estimates and Average Abatement Cost of IRA Tax Credits in the Power Sector from US-REGEN



Sources: Environmental Protection Agency (2022); Rennert and others (2022); and authors' calculations.

Note: Social cost of CO_2 values come from EPA (2022) and Rennert and others (2022), which reflect damages only from sea-level rise, building energy expenditures, temperature-related mortality, and agriculture. Scenario results from US-REGEN are summarized in table 2.

Table 2. Comparison of Key Metrics across Lower, Central, and Higher Fiscal Cost Sensitivities in US-REGEN

Metric	*Lower*	*Central*	*Higher*
Total tax credit expenditures to 2031 ($ billion)	$244	$781	$1,070
Expenditures relative to CBO/JCT	0.9	2.9	4.0
Total fiscal costs to 2031 ($ billion)	$365	$902	$1,190
Economy-wide CO_2 in 2030 (% from 2005)	–28%	–35%	–40%
Power sector CO_2 in 2030 (% from 2005)	–51%	–64%	–75%
Abatement cost ($/t$CO_2$)	$61–87	$45–61	$36–49
Electric vehicle sales share in 2030 (%)	33%	44%	60%
Clean capacity additions to 2035 (GW/year)	34.7	51.3	75.9

Source: Authors' calculations.

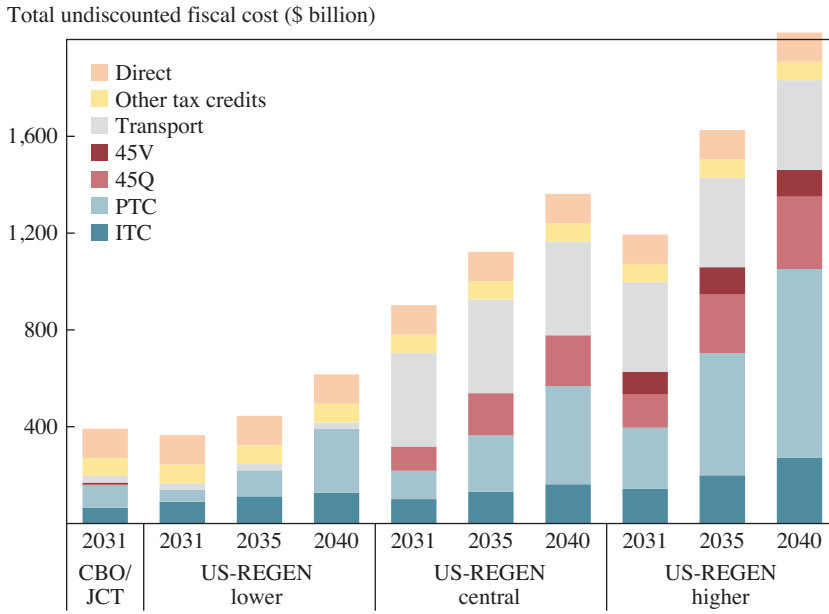
Note: Total fiscal costs include all tax credits and direct expenditures. Abatement costs show ranges across different discount rates (1.5 percent to 3 percent).

assumes higher supply-side costs, higher interest rates, and lower eligibility of electric vehicles for credits. Scenario assumptions are discussed in section VII, which provides a deep dive into how the assumed macroeconomic environment can alter the effects of the IRA. A *higher fiscal costs* scenario uses the same technology and market assumptions as the central case but adds additional policies, including higher credit values for clean hydrogen under the IRA, IRA bonus credits for energy communities and domestic content for all power sector projects, and new and existing source performance standards for power plants.[27]

These sensitivities expand the range of tax expenditures through 2031 from $781 billion in the central case to $244 billion to $1,070 billion, which is similar to the tax credits CBO/JCT score at the low end of the range and four times higher than the CBO/JCT estimate at the high end (table 2). Total fiscal costs of the IRA's climate provisions, inclusive of tax credits and direct expenditures, range from $365 billion to $1,190 billion through 2031 (compared with $902 billion in the central case). This sensitivity with lower fiscal costs also leads to more limited emissions reductions (similar to reference levels in figure 5) because of lower electric vehicle and clean electricity capacity deployment. The higher fiscal costs sensitivity reaches

27. IRA hydrogen credits are assumed to provide $3/kg$H_2$ for all electrolytic hydrogen in this scenario (unlike the central scenario, which assumes more stringent implementation guidance). The central case assumes that only a fraction of projects receive the energy communities bonus for the PTC and ITC; this assumption is consistent with updated Treasury and IRS guidelines, though this guidance is not explicitly modeled (White House 2023b). Model implementation of the new and existing source performance standards for power plants under sections 111(b) and 111(d) of the Clean Air Act comes from the analysis in Bistline (2023)—specifically, the scenario with plant-level cofiring-based standards for coal and natural gas combined cycle units.

Figure 7. Estimates of Cumulative (Undiscounted) Fiscal Costs from IRA Tax Credits by Provision across Low, Central, and High Fiscal Cost Sensitivities in US-REGEN



Sources: CBO and authors' calculations.

Note: CBO/JCT scores are based on September 7, 2022, estimates. Values are shown in nominal terms. "Other tax credits" includes additional end-use incentives (e.g., credits for heat pumps) and manufacturing; 45Q credits are for captured CO_2, and 45V credits are for clean hydrogen.

40 percent economy-wide CO_2 reductions by 2030 relative to 2005 levels and 75 percent reductions in power sector CO_2, which entails power sector investments increasing by almost 50 percent over the central case.

Figure 7 shows IRA fiscal costs by provision across these three sensitivities. Tax credit expenditures quadrupling through 2031 under the higher fiscal costs scenario are due to higher power sector credit uptake vis-à-vis the central case and about $100 billion in hydrogen credits. Fiscal costs of these tax credits approach $2 trillion through 2040 in the higher scenario.

II.F. Caveats

There are several uncertainties associated with the fiscal implications of the IRA. First, specific guidelines about IRA provisions still await clarifications from Treasury and the IRS, including those about bonus credit eligibility, qualifying resources for technology-neutral PTC and ITC in the power

sector, emissions accounting for several credits, and others. Second, forthcoming policy and regulatory changes may affect uptake of tax credits, especially since the timing and stringency of such policies could be influenced by the presence of IRA incentives. For instance, the EPA released proposed performance standards for new and existing power plants under sections 111(b) and 111(d) of the Clean Air Act in May 2023. Since these standards are based on carbon capture and hydrogen at coal- and gas-fired plants, increased deployment of these IRA-subsidized resources could increase and accelerate fiscal impacts (Bistline 2023). Third, there are general uncertainties, as with any projection. For example, US-REGEN assumes that the levelized cost of electricity from utility-scale solar will decline by about 44 percent with the IRA, relative to current unsubsidized levels (see figure 8), though there is considerable uncertainty about the pace of future technological change (Bistline and others 2022; Way and others 2022). These unknowns about technologies, markets, and policies mean that IRA incentive uptake (shown in figure 4) and budgetary impacts are uncertain.

III. Energy Production Analysis

This section describes results from the US-REGEN model (described in section II) that shed light on costs and market implications of the IRA. Our focus here is on the impact of the IRA on the cost of different generating technologies and implications for the price of electricity.

III.A. Levelized Cost of Electricity

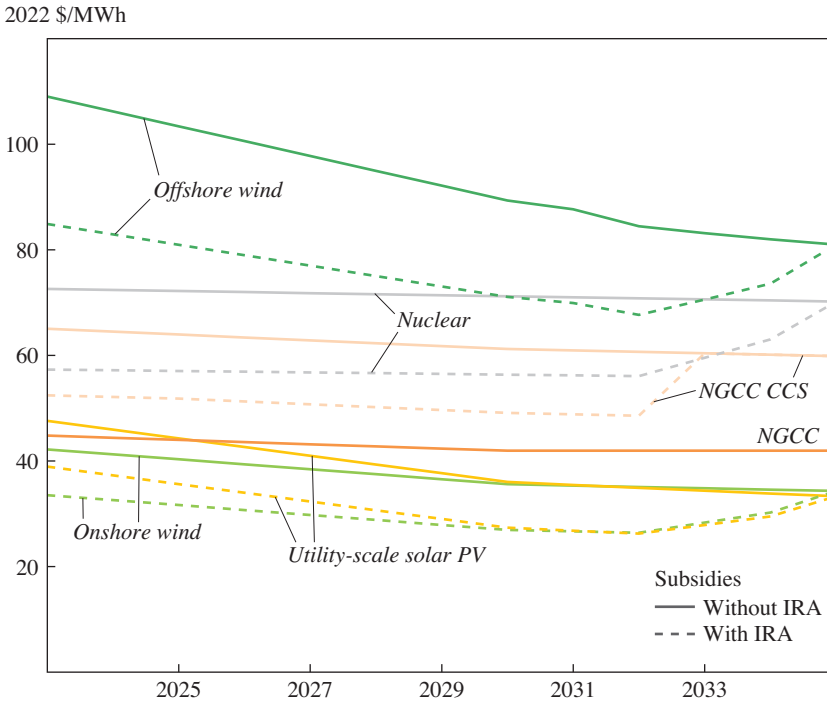
Analysts often reference the levelized cost of electricity (LCOE) for different types of electricity generation. The LCOE is the discounted sum of costs associated with building and operating a power plant over its lifetime divided by the discounted sum of future electricity production:

$$LCOE = \frac{\sum_{t=1}^{T} \left(I_t + OM_t + FC_t \right) \Big/ \left(1 + r \right)^t}{\sum_{t=1}^{T} E_t \Big/ \left(1 + r \right)^t}, \tag{1}$$

where I is investments in period t (including financing costs, minus subsidies), OM is operations and maintenance costs, FC is fuel costs, r is the discount rate, and E is electricity production.

Figure 8 plots estimates of the levelized costs of electricity through 2035 for several prominent generating technologies. The estimates are

Figure 8. Levelized Cost of Electricity by Technology over Time without and with IRA Subsidies



Sources: Bistline, Roney, and others (2023) and authors' calculations.

Note: Capital costs underlying these estimates are based on Bistline, Roney, and others (2023). Estimates assume thirty-year financial lifetimes and 7 percent discount rates (weighted average cost of capital). Onshore wind and utility-scale solar PV are assumed to take the PTC under the IRA, nuclear and offshore wind take the ITC, and natural gas combined cycle with CCS (NGCC CCS) takes CO_2 capture credits. Labor bonuses are included but not bonuses for energy communities or domestic content.

based on capital cost assumptions from the EPRI's US-REGEN model used for the analysis in other sections.[28] Without the IRA, costs for renewable technologies, including solar and wind, are generally projected to decline (24 percent for solar, 16 percent for onshore wind, and 18 percent for offshore wind by 2030). These projections are based on a combination of factors, including assumptions about learning curves, technological

28. Values are discussed in Bistline, Roney, and others (2023). Capital cost assumptions in 2030 are similar to the National Renewable Energy Laboratory's Annual Technology Baseline (moderate technology innovation scenario), which is the primary source for many models that have informed policy discussions about the impacts of the IRA (see Bistline, Blanford, and others 2023); however, near-term costs have been adjusted to reflect observed 2022 costs.

progress (e.g., larger rotors for onshore wind), and estimates from the literature.[29]

Figure 8 also highlights the impact of IRA subsidies on the estimated costs for all the eligible technologies. Only the natural gas combined cycle without CCS is ineligible for subsidies. The calculations behind figure 8 assume that the technology-neutral production and investment tax credits begin to phase down in 2032 for illustrative purposes, though these could continue afterward if power sector CO_2 emissions have not reached 25 percent of their 2022 levels.

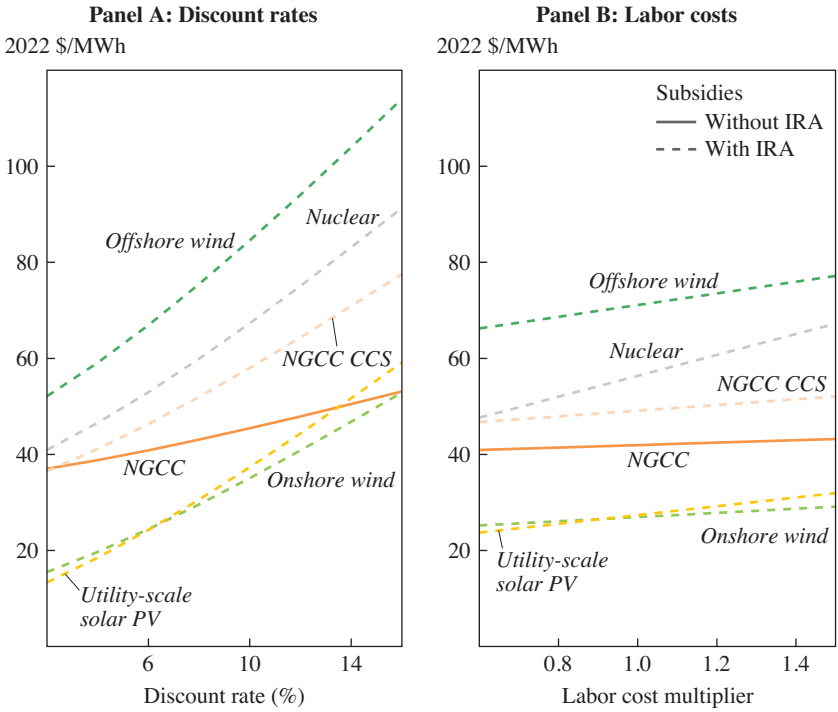
While LCOE estimates can be a useful summary statistic to make high-level comparisons, especially for time trends of individual technologies, they also have well-known limitations. The statistic does not account for the value of the electricity provided to the grid. LCOE does not distinguish between resources that produce electricity primarily when demand and wholesale prices are low and resources whose output aligns with higher-priced periods, including dispatchable resources. Joskow (2011) details these issues, highlighting the extent of the problem by noting that "the difference between the high and the low hourly prices over the course of a typical year, including capacity payments for generating capacity available to supply power during critical peak hours, can be up to four orders of magnitude" (239). The value to production at different times of the day and over the course of the year will also vary as more renewable resources are added to the system.[30] Models that minimize system costs (such as US-REGEN) simultaneously account for power system investments and operational dynamics, including endogenous representations of how the costs and value of different resources change as the system evolves (Bistline, Blanford, and others 2021).

With those limitations in mind, figure 8 highlights how, particularly with IRA tax credits, the LCOE of onshore wind and solar PV installations is below that of other resources. Firms will have strong incentives to build these resources. Offshore wind, even with IRA subsidies, is considerably more expensive than other resources, but several states have resource-specific

29. This figure does not show variability in LCOE estimates based on regional labor costs, resource quality (e.g., for wind and solar generation), or capacity factors (which can vary by scenario and time period), though these features are accounted for in the US-REGEN modeling in other sections.

30. In particular, wind, solar, and other resources exhibit diminishing marginal returns, where the economic value of additional capacity decreases as their deployment increases (Bistline 2017; Hirth 2013). Metrics like LCOE neglect declining value and increasing system costs.

Figure 9. Levelized Cost of Electricity by Technology in 2030 for Various Assumptions about Discount Rates and Labor Costs



Source: Authors' calculations.

Note: All other parameters are held constant from the comparisons in figure 8. Costs for technologies are shown after accounting for IRA subsidies, except for NGCC capacity, which is not eligible for tax credits.

mandates for these resources, which are represented in many models, including US-REGEN. Finally, the impact of tax credits on the LCOE is a function of the assumptions about how long the subsidies will persist. As discussed above, we illustrate LCOEs with credits that begin declining in 2032, but extended tax credits would reduce LCOEs across longer horizons, potentially across multiple decades.

LCOE estimates, including those reflected in figure 8, incorporate assumptions about the interest rate faced by a project developer. Figure 9, panel A, plots LCOEs under different discount rates, highlighting how various technologies respond to increases in the cost of borrowing (across an illustrative range of rates). Low-carbon technologies are all more sensitive to increases in interest rates, reflecting the large upfront investments

required and relatively low operating costs, including a lack of fuel costs (except in the case of CCS-equipped capacity). Figure 9, panel B, plots LCOEs under different assumptions about labor costs. Nuclear plants are the most sensitive to increases in labor inputs, and labor's share of total plant costs ranges from 17 percent (for onshore wind) to 44 percent (nuclear). Section VII below explores the sensitivity of IRA fiscal costs and emissions impacts under different assumptions about interest rates and other input costs.

III.B. Electricity Market and Price Impacts of Tax Credits

In addition to encouraging the construction of new electricity generating resources, IRA tax credits will have an impact on how new and existing resources are operated. IRA incentives can have large impacts on electricity markets, since credits can lower wholesale prices and increase the prevalence of negative-priced periods. Generation technologies that collect production-based tax credits will have strong incentives to operate even when wholesale prices are low and even negative (which is more common during hours when output from these resources is highest) to collect credits.[31] For example, the variable costs of operating a wind turbine are negligible, and the operator could receive as much as $33/MWh in subsidies. As long as wholesale prices are above –$33/MWh, it is profitable for the wind plant to generate electricity.[32] Negative prices alter economic signals for market entry and exit of generators, shift incentives for locational decisions and balancing resources (e.g., energy storage, transmission), affect system operations, and change the economics of end-use electrification and new loads (e.g., hydrogen production, cryptocurrency mining).

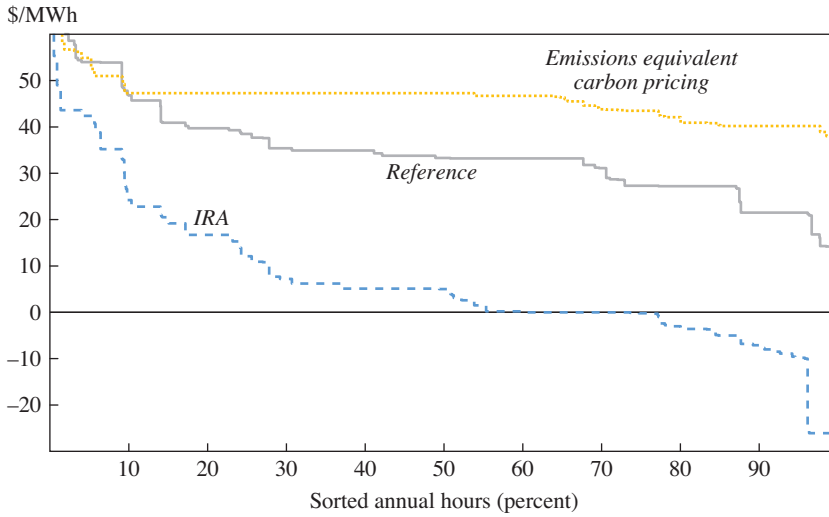
Electricity markets have already experienced periods of negative wholesale prices, driven in large part by subsidized renewables.[33] For example, California and Texas, areas with significant renewable capacity, experienced the most periods with negative wholesale electricity prices in 2021 and the first half of 2022 (Malik 2022). Energy storage can play an important role in shifting electricity from periods with low prices to periods with higher prices. Since the IRA extends tax credits to stand-alone energy

31. Output-based tax credits under the IRA include the technology-neutral PTC, credits for captured CO_2, and credits for clean hydrogen.

32. These negative offer prices can be as low as –$70–90/MWh for coal with CCS taking 45Q credits for captured CO_2 (Bistline, Roney, and others 2023).

33. Negative bids from subsidized resources are not the only cause of negative prices in current markets. Local transmission congestion and system-wide oversupply (e.g., due to flexibility constraints) also play roles (Seel and others 2021).

Figure 10. Wholesale Electricity Price Duration Curves for the Reference, IRA, and Carbon Price Scenarios



Source: Authors' calculations.

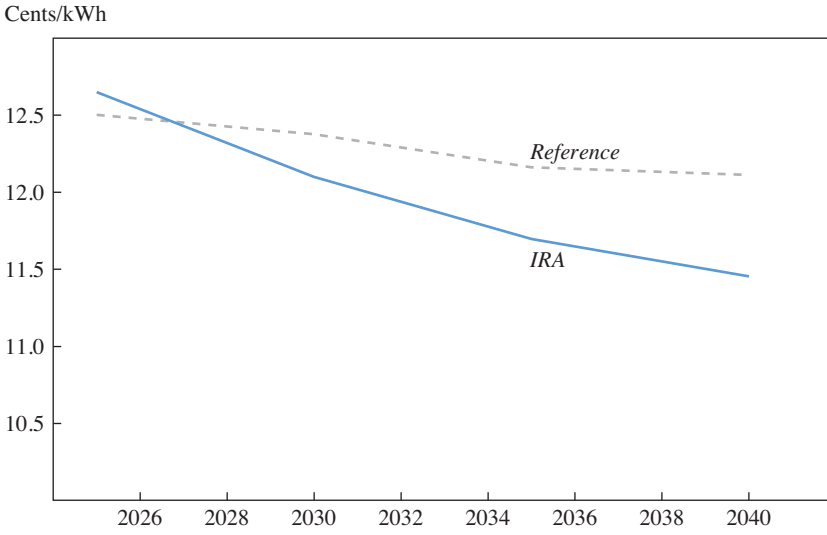
Note: Curves are shown for the Southwest Power Pool (SPP) region in 2050, which includes South Dakota, Nebraska, Kansas, and Oklahoma.

storage, deployment of these resources could be significant: comparisons in section II illustrate how deployment of storage technologies could total over 10 GW per year (compared with about 7 GW of energy storage installed cumulatively as of 2022).

Even with storage deployment, figure 10 shows how IRA tax credits increase the frequency of zero- and negative-priced hours, which comprise nearly half of all hours in the wind-dominant Southwest Power Pool (SPP) region. Ultimately, the frequency of negative-priced periods depends on (1) the fraction of generators taking different tax credits (since negative-priced periods are more likely the more generators take the PTC rather than the ITC), (2) the extent of generator entry and exit (which affects shares of IRA-subsidized resources), and (3) regional supply-demand balances (which influence supply curves and ultimately which technologies are on the margin).[34] In contrast, emissions equivalent carbon pricing (as presented

34. There are many hours with positive clearing prices, even with significant deployment of zero-marginal-cost resources—energy storage bids at its opportunity cost, gas-fired capacity is on the margin in many hours, and not all subsidized resources are incentivized to make negative bids (e.g., those electing to take the ITC).

Figure 11. Load-Weighted National Average of Residential Retail Electricity Prices over Time



Source: Authors' calculations.
Note: Values are based on US-REGEN modeled scenarios with IRA incentives and a counterfactual reference without IRA. Note the truncated vertical axis.

in table 3) increases wholesale prices relative to the reference. In terms of policy, there are different perspectives on the importance of wholesale and retail market changes for deeply decarbonized energy systems, which may be dominated by resources with zero or negative short-run marginal costs, energy-limited devices such as storage, and cross-sector interactions and subsidized resources, which may exacerbate "missing money" problems and out-of-market payments for resource adequacy and reliability (Ela and others 2021; Mays, Morton, and O'Neill 2019; Hogan 2019; Conejo and Sioshansi 2018).

While the analysis of price pressures has focused on the production-based incentives in the electricity market, similar incentives will exist in the manufacturing sector, where the IRA offers subsidies per unit produced (e.g., solar panels). These subsidies will put downward pressure on market prices for all suppliers.

Annual retail electricity price impacts of IRA incentives are shown in figure 11. Modeled prices in US-REGEN illustrate how IRA incentives lower national long-run retail prices by 2.2 percent in 2030 and 5.4 percent in 2040, relative to a counterfactual scenario without IRA incentives,

which already indicates a trend of declining prices over time.[35] These changes are consistent with other studies of the IRA (Roy, Burtraw, and Rennert 2022), which also depend critically on assumed fuel price trajectories. Beyond lowering household and firm electricity costs, an advantage of retail electricity price declines is that they can further encourage end-use electrification, which is a central decarbonization strategy in many studies (DeAngelo and others 2021).

III.C. Distributional Implications

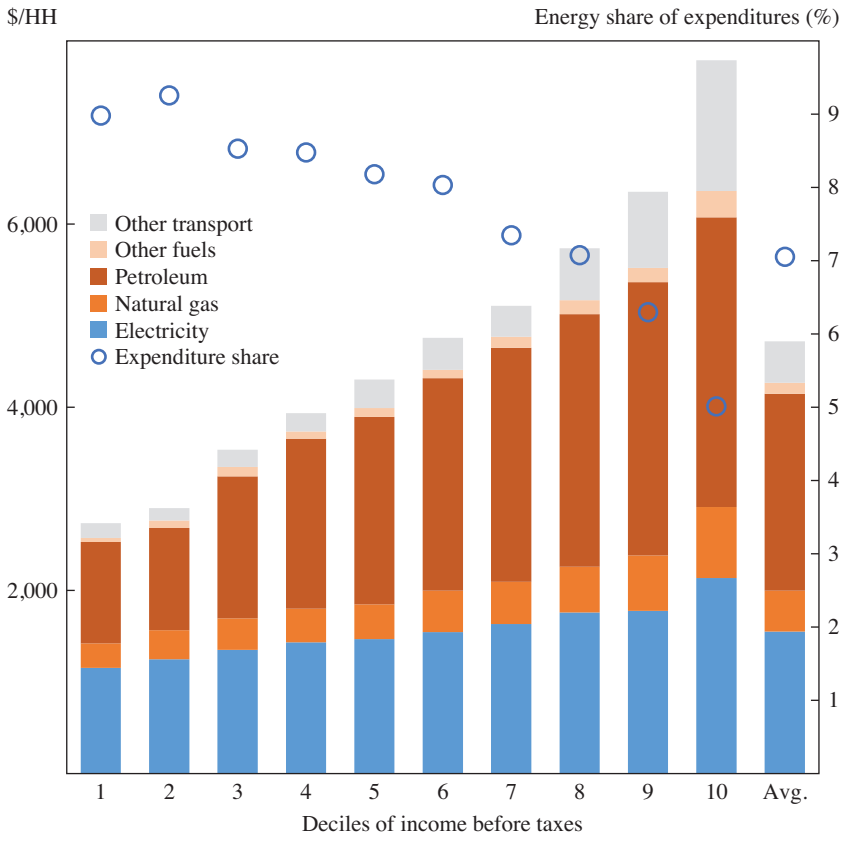
Households across the income distribution will experience different impacts from the IRA. Lower-income households spend a slightly larger share of their income on electricity than higher income households (figure 12), so by reducing the costs of electricity, the IRA will provide greater benefits for many lower-income households. The relationship between income and consumption shares is stronger for petroleum (primarily gasoline for transport), so the means-tested tax credits for electric vehicles in the IRA can help lower-income households that buy an electric vehicle reduce these expenditures. For example, Burnham and others (2021) find that even at gasoline prices of less than $2.63 per gallon, fuel costs for electric vehicles are about half the fuel costs for a comparably sized vehicle with an internal combustion engine. Impacts over the next decade or more will depend heavily on expected electricity and gasoline prices. Burnham and others (2021) also suggest that maintenance costs are significantly lower for electric vehicles. Finally, Linn (2022) describes how electric vehicle subsidies may lead to lower prices for gas-powered vehicles by interacting with existing fuel economy standards and state-level zero-emission vehicle programs. Since low-income households are more likely to buy gas-powered vehicles, this effect is progressive.[36]

The economic incidence of the tax credits will also factor into the distributional outcomes. US-REGEN assumes perfectly competitive markets, meaning that the tax credits are passed through to consumers. In electricity

35. Retail prices in US-REGEN are built up from modeled generation prices. These regional prices are calibrated to observed base year prices, including a markup to reflect sunk costs built into the rate base. Future retail markups are scaled based on projected changes to transmission and distribution costs as a function of the changing load and resource mix. Note that prices increase in 2025 under the IRA due to the assumption in the reference case that solar and wind tax credits maintained their previous step-down schedule, which leads forward-looking firms to front-load investment in these resources to take advantage of expiring credits.

36. US-REGEN captures many of these effects, and in ongoing work, we quantify the expected impacts of the IRA by region and income level.

Figure 12. Annual Household Energy Expenditures across Income Deciles



Source: US Bureau of Labor Statistics.

Note: Data are from the BLS 2022 Consumer Expenditure Surveys, table 1110, showing data for 2021 for deciles of income before taxes; https://www.bls.gov/cex/tables/calendar-year/mean-item-share-average-standard-error/cu-income-deciles-before-taxes-2021.pdf. Energy shares of expenditures on the secondary axis are based on average annual expenditures for consumer units within the decile.

markets, this assumption means that wholesale electricity buyers pass on the full benefits of the subsidies to end-use consumers. The second step involves assumptions about the political and regulatory processes that determine regional retail rates. Most other IRA analyses make the same assumption that the full values of the tax credits are passed through to consumers.[37] In reality, electric vehicle manufacturers, clean electricity

37. The analysis by Cole and others (2023), which presents several scenarios that vary the share accruing to the consumer, is the lone exception of which we are aware.

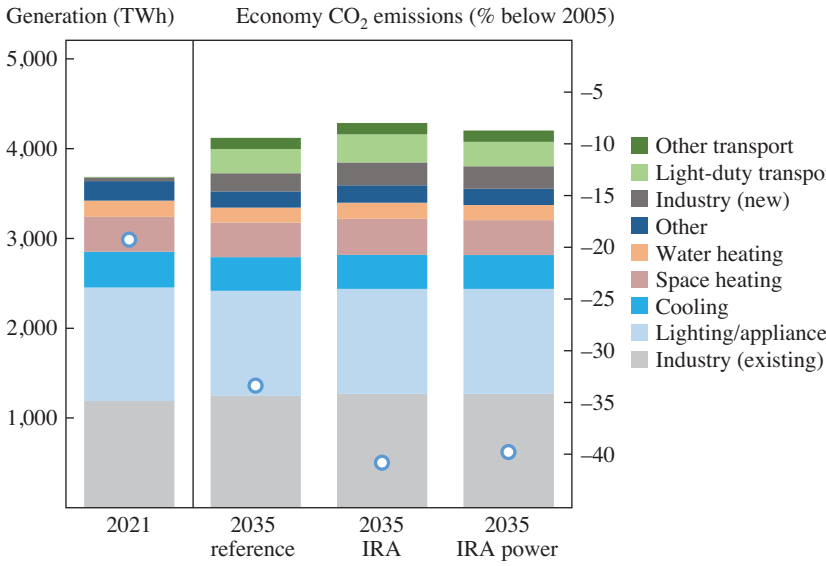
producers, and other firms and their shareholders may capture some of the tax credits. As described in online appendix B, an index of clean energy stocks fell relative to the market on December 19, 2021, when Senator Manchin, the pivotal fiftieth Democratic vote, went on television to say he was done negotiating on the bill then known as Build Back Better, which contained many of the same climate provisions as the IRA. It increased on July 27, 2022, when the Senate deal on the IRA was announced. These movements suggest that producers could gain from the tax credits and that they will not be fully passed through to consumers. Particularly if the domestic content and labor provisions change behavior (i.e., they are marginal), non-energy firms and their workers may also gain from the tax credits. In the case of electric vehicles, the incidence of the credits, and perhaps even the market structure for vehicle sales, will depend crucially on whether the Treasury guidance exempting leased vehicles from requirements on battery sourcing and eligibility caps stands. Sorting out the incidence of different IRA provisions and measuring it empirically will be useful tasks for future research.

III.D. Electrification Implications

The IRA alters electricity demand through two channels—by directly subsidizing the adoption of electric end-use technologies (e.g., the electric vehicle and heat pump tax credits discussed in section I.A) and by lowering electricity prices (as discussed in the section III.C). Technological change and consumer choice lead to transport, industrial, and buildings electrification and to a 12 percent increase in electricity demand by 2035 in the reference scenario without the IRA (figure 13). IRA incentives increase load growth by 4 percentage points to 16 percent from current levels by 2035.

To isolate the impact of lower prices on electricity demand, we include a scenario with only the IRA's power sector tax credits, which reach prices similar to those shown in figure 10. By 2035, demand for electricity increases in this scenario by 14 percent from current levels, which is midway between the reference without the IRA (12 percent) and the IRA scenario (16 percent), as shown in figure 13, with most of these increases relative to the reference coming from industry. Since there are more-limited IRA tax credits that directly incentivize industrial electrification, electricity growth in industry under the IRA is attributed primarily to lower electricity prices, which are reflected in high industrial electricity demand even with power sector IRA credits only (96 percent of growth in industrial demand under the IRA from the reference persists with only power sector credits).

Figure 13. Electricity Demand by End Use



Source: Authors' calculations.

Note: Values for 2035 are based on US-REGEN scenarios without IRA (reference), one with all IRA incentives (IRA), and one with power sector IRA incentives only (IRA power).

In contrast, incremental passenger vehicle electrification under the IRA is largely from the clean vehicle credits, as less than 2 percent of growth in passenger vehicle electricity demand from the reference occurs with power sector provisions alone. There are several reasons for this muted demand elasticity, even with lower wholesale electricity prices (see figure 10), including stock turnover dynamics, limited effects of fuel costs in purchase decisions, and wholesale electricity prices being only one component of retail prices (i.e., unsubsidized transmission and distribution costs could comprise large shares of retail prices).

IV. Macroeconomic Framework

In this section, we consider the macroeconomic impacts of the climate provisions of the IRA. We present a conceptual framework for understanding the macroeconomic impacts of clean power tax credits. Using a neoclassical growth model with clean energy capital, we show both the long-run and short-run macroeconomic impacts of investment and production tax

credits on macroeconomic variables: interest rates, wages, output, and consumption. We show how macroeconomic outcomes vary depending on labor market conditions, bottlenecks in clean energy investment, reductions in the price of capital due to learning by doing, and domestic sourcing requirements. The model details are presented in online appendix A.

IV.A. Impact of Tax Credits

We start by characterizing how increases in clean energy tax credits impact macroeconomic aggregates. For simplicity, assume that household electricity demand is inelastic and fixed at $\overline{E}^h$. Then changes in power generation impact electricity prices and demand only via industrial demand. Model equilibrium can be reduced to the following three equations:

$$(2) \qquad p_t^c\left(1-\tau_t^{inv}\right)u_c\left(C_t\right)=\beta u_c\left(C_{t+1}\right)\left[\begin{array}{c}\left(F_e\left(E_{t+1}^f,\overline{N}\right)+\tau_{t+1}^p\right)G_c\left(K_{t+1}^c\right)\\ +\,p_{t+1}^c\left(1-\tau_{t+1}^{inv}\right)\left(1-\delta_c\right)\end{array}\right],$$

$$(3) \qquad C_t=F\left(E_{t+1}^f,\overline{N}\right)-p_t^c\left(K_{t+1}^c-\left(1-\delta_c\right)K_t^c\right), \text{ and}$$

$$(4) \qquad G\left(K_t^c\right)=E_t^f+\overline{E}^h,$$

which jointly determine the equilibrium path of electricity supplied to industry E_t^f, household consumption C_t, and clean energy power generation K_{t+1}^c as a function of underlying parameters and the exogenous path of capital prices p_t^c and fiscal policy.

In steady state, the level of clean energy capital is defined implicitly by the following condition:

$$(5) \qquad p_c\left(\frac{1}{\beta}-1+\delta_c\right)\left(1-\tau_{inv}\right)=\left(F_e\left(G\left(K_{ss}^c\right),\overline{N}\right)+\tau_p\right)G_c\left(K_{ss}^c\right).$$

The left-hand side of this equation is just the steady-state user cost of capital, while the right-hand side is the marginal product of capital taking into account the effect of both power generation capital on electricity

production and electricity production on overall output. Under mild conditions for the production function F and the power generation function G (see online appendix A), the steady-state level of capital K_{ss}^{c} is increasing in both the investment tax credit and the production tax credit. Figure 14 shows the user cost of capital and the marginal product of capital, which pin down the steady-state level of clean energy capital. The shift in user cost (left-hand side) and marginal product (right-hand side) under an investment and production tax credit are shown.

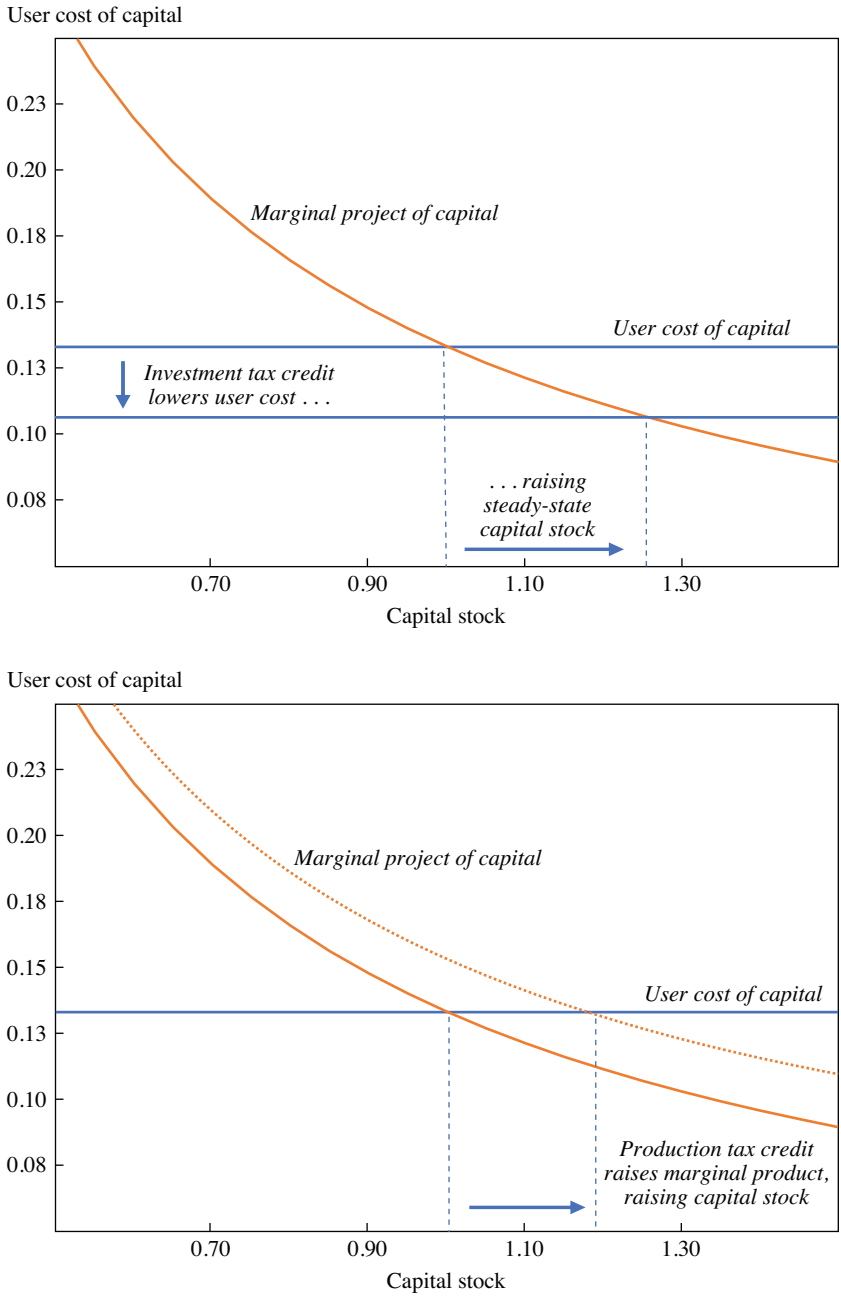
An increase in the clean energy capital stock increases other macroeconomic quantities in steady state; the increase in electricity supplied to industry raises output and investment. Labor productivity and wages also rise due to an increase in industrial electricity supply. Assuming that the subsidies are not too large, consumption also increases in steady state. On the price side, the price of electricity falls, while the real interest rate is unchanged ($r = 1/\beta - 1$).

Effectively, production and investment subsidies are a negative capital tax, raising the capital stock and productivity. To the extent that the weight of energy in the aggregate production function is low or the marginal product falls sharply with higher electricity supply, the tax credits have only marginal benefits in terms of productivity, output, wages, and consumption. However, the price of electricity would fall more sharply in this case. As with other supply-side tax policies with dynamic effects, the fiscal burden would be tempered somewhat by higher output (relative to a static analysis).

The transition path to this new steady state can also be characterized. An increase in the investment or production tax credit lowers consumption on impact while raising investment. The real interest rate rises initially before gradually falling back to its long-run level. Labor productivity, wages, and electricity prices inherit the dynamics of the capital stock, with wages and productivity rising gradually along the transition path to their higher steady state level. On the fiscal side, the increase in interest rates would raise debt-servicing costs on impact, with higher output mitigating the impact on the debt-to-GDP ratio over time.

Given that output, consumption, investment, wages, and productivity ultimately rise over the long run (and the price of electricity falls), a natural question to ask is if a strictly positive subsidy is welfare-improving. So far, we have not considered the impact of the policy on emissions or in mitigating damages from climate change, so the question is solely about whether the macroeconomic benefits are welfare-improving. The answer is no.

Figure 14. Steady-State Capital after an Increase in Investment Tax Credit and in Production Tax Credit



Source: Authors' calculations.

For the planner, the aggregate resource constraint holds, and the optimal allocation for clean energy capital would be implemented only if both the ITC and PTC are set to zero.

IV.B. Alternative Macroeconomic Conditions

In the baseline model, we assumed a fixed, inelastic supply of labor. Supporters of the Inflation Reduction Act have argued that by subsidizing construction and manufacturing, the legislation may help create jobs by increasing the labor force and by creating jobs in communities/regions that have seen structurally weaker labor markets (Williams, Zhavoronkov, and Madland 2022). Power generation would likely stimulate labor demand in less urbanized regions, and upstream labor demand for raw materials and equipment are also likely to be present in regions that have seen less robust labor markets.

One way to incorporate this channel is to assume that there exists a pool of labor that is currently outside the labor force but has a relatively low reservation wage. If aggregate labor supply is given by an upward-sloping labor supply curve (i.e., $w_t = N_t^{\varphi}$), then the investment or production tax credit increases employment as a higher capital stock boosts the marginal product of labor and labor supply adjusts to meet higher labor demand. In steady state, the increase in the clean energy capital stock is larger than in the case of an inelastic labor supply. Steady-state consumption, output, and labor productivity all increase by more than in the baseline case. This case also sees a larger level of subsidies, but a sharper decline in the price of electricity and possibly a decline in subsidy cost as a share of GDP.

However, a more elastic labor supply results in a *larger* increase in interest rates and decline in consumption (relative to the baseline case) as the initial demand impact dominates; an elastic labor supply implies a larger increase in investment given a larger increase in the desired capital stock. Along the transition path, as the capital stock rises, employment rises to keep the marginal product of labor from rising as much as in the baseline case, and the real interest rate gradually converges back to its long-run value $1/\beta - 1$.

Given supply chain disruptions and dislocations during the pandemic, another relevant departure to consider is the presence of bottlenecks that constrain the ability to ramp up construction of clean energy facilities or the manufacturing of equipment for those facilities. Domestic content requirements or delays in siting or transmission access could function to either raise the cost or outright prevent investment in new clean power generation despite the financial incentives provided in the IRA. Bottlenecks

are classified as either "market" or "nonmarket"; the former shows up as an increase in the relative price of capital p_t^c when a (possibly time-varying) investment constraint is reached, while the latter is simply a constraint on investment that does not impact price. Market bottlenecks could be thought of as inelastic supply of a key material (like lithium or copper), while nonmarket bottlenecks could be thought of as prohibitions on the siting of new solar or wind projects.

Our model can be generalized to account for nonmarket bottlenecks by simply introducing a constraint on investment: $I_t^c \leq \overline{I}$. The intertemporal optimality condition for investment is now

$$p_t^c\left(1-\tau_t^{inv}\right)\lambda_t+\mu_t=\beta\lambda_{t+1}\left[\begin{array}{c}\left(F_e\left(E_{t+1}^f,\overline{N}\right)+\tau_{t+1}^p\right)G_c\left(K_{t+1}^c\right)\\ +\,p_{t+1}^c\left(1-\tau_{t+1}^{inv}\right)\left(1-\delta_c\right)\end{array}\right] \tag{6}$$
$$+\,\beta\mu_{t+1}\left(1-\delta_c\right),$$

where μ_t is the Lagrange multiplier on the investment constraint in period t.

If the constraint is binding in steady state, then investment is constrained at its upper bound $\overline{I}$, and the user cost of capital is above the level that would obtain absent any bottlenecks. Relative to the baseline case, the capital stock is lower, as is output. The price of electricity is higher, and subsidies paid are lower. Even when the constraint does not bind in steady state, bottlenecks may constrain investment along the transition path. In this case, the path of investment is below the baseline case until the desired investment rate falls below the constraint. Since investment is lower initially, the rise in the interest rate and decline in consumption are muted in the case where bottlenecks initially bind. Bottlenecks mitigate the macroeconomic impact since the rise in investment is attenuated and the required decline in consumption to meet desired investment is lessened.

Market bottlenecks that manifest as a higher price of investment carry the same implications for the steady-state capital stock and transition dynamics. Indeed, the Lagrange multiplier in the nonmarket bottlenecks case is effectively a shadow price. For concreteness, assume that $p_t^c = p_c$ if $I_t^c < \overline{I}$—that is, the price of investment is constant at its steady-state value p_c so long as the investment constraint is slack. Then we can define a market price p_t^c when $I_t^c = \overline{I}$ that satisfies the investment Euler equation above:

$$p_t^c = p_c + \frac{\mu_t}{\lambda_t\left(1-\tau_t^{inv}\right)}. \tag{7}$$

Substituting for μ_t in the Euler equation above returns a Euler equation that delivers the same equilibrium path of consumption and investment as the nonmarket bottlenecks case. Likewise, the steady-state level of the capital stock would also be depressed (relative to baseline) because the tax credits are offset by a rise in the user cost of capital.

The distinction between market and nonmarket bottlenecks is not relevant for investment, but it does carry implications for the fiscal cost. With both types of bottlenecks, less clean energy investment is deployed, lowering fiscal cost. But market bottlenecks carry a counteracting effect on prices. Despite lower real investment, a higher price of investment may be sufficient to raise the fiscal cost of an investment tax credit. This issue is unique to the ITC since the PTC is tied to the quantity of clean energy produced. In the extreme case that investment is already against its constraint, the tax credit is fully offset by a rise in the price p_t^c, and the government subsidy generates no new investment. Indeed, there may a Laffer curve for ITC fiscal cost where fiscal cost is non-monotonic in the level of investment; with extreme bottlenecks no investment occurs, so fiscal expenditure is zero, but very high prices due to bottlenecks keep fiscal expenditure high even for low levels of real investment.

IV.C. Learning by Doing

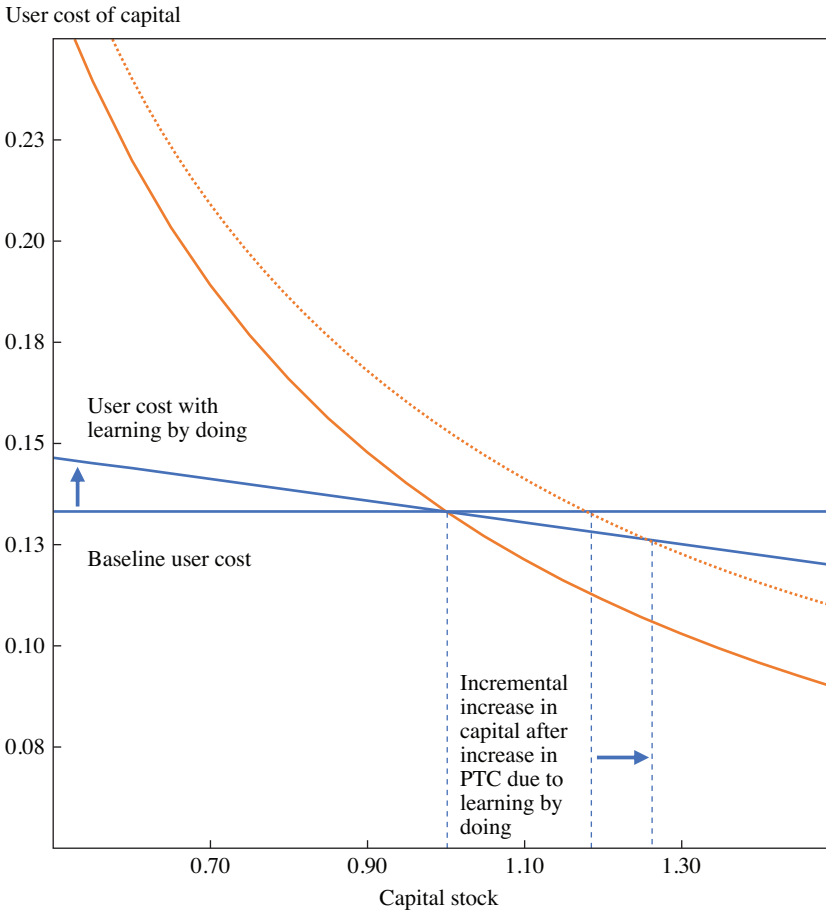
Proponents of the subsidies approach to clean energy transition have emphasized the dramatic decline in solar and wind production costs over the last decade and the potential for further declines in clean energy production costs. The Inflation Reduction Act includes a host of incentives to support the supply chain for clean energy and generous incentives for technologies that are not yet cost competitive, like carbon capture and clean fuels. To the extent that higher production spurs learning by doing and cost reductions, how might that impact the clean energy transition?

A classic formulation of learning by doing is Wright's law, which expresses marginal cost as a function of cumulative production (Wright 1936; Way and others 2022).[38] Accordingly, we model learning by doing by making the price of capital a decreasing function of the current stock of clean energy capital: $p_t^c = p_c\,(K_t^c)$.[39] For modest levels of learning by doing, the user cost of capital falls as more clean energy capital is produced.

38. The presence of credit constraints or locally increasing returns to scale would generate similar results.

39. Modulo of depreciation, which is low for power structures, the current capital stock is equal to cumulative production.

Figure 15. Learning by Doing Results in Larger Increases in Steady-State Capital after an Increase in PTC



Source: Authors' calculations.

Relative to the baseline model, this implies a larger increase in the clean energy capital stock in steady state for a given subsidy policy. Clean energy subsidies now lead to a larger increase in output, labor productivity, and wages and a larger decline in the price of electricity.[40] Figure 15 shows how the user cost becomes downward sloping with learning by doing, magnifying the impact of either the ITC or PTC.

40. If the price of capital is sufficiently sensitive to production, multiple steady states are possible. In particular, if over some region the price of capital drops faster than the marginal product of electricity, multiple steady states will obtain.

IV.D. Domestic Sourcing

As noted earlier, the IRA's investment tax credit is eligible for bonuses based on whether labor standards and domestic sourcing requirements are satisfied. The labor and domestic sourcing requirements have two distinct impacts on the equilibrium level of investment that is eventually realized. One way to model domestic sourcing is to assume two distinct prices for investment: $p_t^{c,low}$ and $p_t^{c,high}$, with associated investment tax credits of $\tau_t^{c,low}$ and $\tau_t^{c,high}$. The household's budget constraint becomes

$$(8) \qquad C_t + p_t^{c,low}\left(1-\tau_t^{c,low}\right)I_t^{c,low} + p_t^{c,high}\left(1-\tau_t^{c,high}\right)I_t^{c,high} = F\left(E_t, \overline{N}\right).$$

In steady state, the household chooses whichever option delivers the lowest user cost of capital. Importantly, the higher relative cost of investment may be chosen if the labor and domestic sourcing bonus is sufficiently generous. In this sense, the labor and domestic sourcing bonuses raise the clean energy capital stock, since the lowest user cost option (inclusive of the tax credit) is chosen in equilibrium. That the domestic bonus may lead to higher clean energy capital is a function of the generosity of the tax credit and does not mean it is optimal. Absent some other benefit or externality, the planner would always choose the lower cost investment $p_t^{c,low}$ to minimize the cost of achieving some given level of emissions.

Domestic sourcing has another distinct element—the use of imports to meet increased clean energy investment demand. Suppose that domestic and foreign producers both sell at a relative (domestic) price of p_t^c and are eligible for the same ITC. Then the absence of domestic sourcing would allow for a faster increase in investment with less crowding out of domestic consumption. The steady-state level of clean energy capital would be unchanged (relative to baseline), but the real interest rate would rise by less, and decarbonization of electricity production would be faster. From an emissions perspective and given conditions of full employment, the inability to import and slowdown in investment are likely to be the most important impact of domestic sourcing requirements.

IV.E. Crowding Out and Capital Taxation

Our model so far has abstracted away from the funding mechanisms of the IRA and possible impacts on the non-energy capital stock. It is straightforward to add non-energy capital to the production function and characterize the effect of the clean energy subsidies on non-energy capital. As was the case before, the subsidies increase the steady-state clean energy capital

stock but also increase the non-energy capital stock. A lower price of electricity increases the rate of return and demand for non-energy capital in the same way that it increases labor demand. There is no long-run crowding out; private non-energy capital is *crowded in*. Output, consumption, wages, and labor productivity all rise in the long run by more than in the baseline model.

However, in the short run, investment in clean energy capital competes with non-energy capital and consumption. Depending on the elasticity of intertemporal substitution, non-energy investment and consumption may fall while clean energy investment crowds out these other uses. For sufficiently large crowding out of non-energy capital investment, it is possible, in the short run, that output, wages, and labor productivity initially decline before rising to their higher steady-state values.

The IRA is estimated to reduce deficits through a combination of reductions in prescription drug spending in Medicare, increased revenue from tax enforcement, and an increase in corporate taxation. So far, we assume that the tax credits in the IRA are funded via a lump sum tax levied on the representative household that does not distort behavior.[41] However, the corporate tax and tax enforcement provisions in the IRA may have an impact closer to a rise in the marginal tax rate on capital. In standard macroeconomic models, corporate tax increases can have substantial effects on investment and the capital stock.

Depending on the magnitude of the capital tax, the steady-state increases in output, wages, and labor productivity may be reversed if the long-run crowding out for non-energy capital is large enough. However, in the short run, the rise in capital taxation would mitigate the crowding out and interest rate effects of the clean energy subsidies. A higher capital tax, by lowering non-energy investment demand, would free output for consumption or clean energy investment, mitigating or reversing the increase in real interest rates. To the extent that the capital taxation effects dominate, output, consumption, and non-energy investment would gradually fall along the transition path to their lower steady-state values.

V. Clean Energy Subsidies versus Carbon Pricing

Subsidy-based approaches and emissions pricing are two widespread policy instruments to reduce emissions, each with trade-offs across economic, environmental, and political dimensions. While the United States does not

41. The macroeconomic impact would be identical if instead the policy were fully deficit financed (assuming that taxes eventually rise to stabilize the debt-to-GDP ratio).

have a federal price on carbon emissions, 55 percent of the greenhouse gas emissions in the rest of the Organisation for Economic Co-operation and Development (OECD) are subject to an explicit Pigouvian carbon price, in the form of either a carbon tax or cap-and-trade system (OECD 2022; Timilsina 2022). Several US states have carbon pricing covering varying shares of their carbon emissions, ranging from under 20 percent in Massachusetts to almost 75 percent in California. In total, explicit carbon pricing covers 6 percent of US GHG emissions.[42]

It is instructive to compare the subsidy-based approach to carbon pricing.[43] At a high level, both subsidies and carbon pricing change the relative prices of non-emitting and emitting fuels. Some of the provisions of the IRA are output-based credits, which directly subsidize production of zero-carbon energy. If every unit of subsidized production offset the same amount of carbon emissions (for example, if all nonsubsidized energy were generated by the same fossil fuel–fired technology with the same emissions per unit of energy), then the production tax credits would be identical to a carbon price. In other words, subsidizing a clean MWh of electricity at $10 per MWh would displace as much carbon from the electricity sector as taxing an emitting MWh of electricity at $10 per MWh, assuming inelastic demand.

V.A. Comparison to a Carbon Tax

To place more structure on the question of subsidies versus carbon pricing, we extend our macroeconomic framework from the previous section to allow for fossil fuel power generation. We now allow for both fossil fuel and clean energy capital, with relative price of investment p_t^f and p_t^c. Electricity generated from fossil fuel capital is given by the generation function $G^f(\bullet)$, with capital as the only factor of production; electricity generated from clean energy capital is given by the generation function $G^c(\bullet)$. Total electricity production is simply the sum of electricity generated from fossil fuel and clean energy capital (i.e., perfect substitutes). The details of the model extension are shown in online appendix A.

Relative to households in the baseline model, households now may invest in both clean energy and fossil fuel capital. Clean energy is eligible for both

42. The World Bank, "Carbon Pricing Dashboard," https://carbonpricingdashboard.worldbank.org/map_data.

43. Most sources describe policies as carbon pricing, although many of the policies cover GHG emissions in addition to CO_2. As noted above, the IRA introduces a price on methane emissions from specified entities in the oil and gas sector. These GHG emissions are not included in the 6 percent figure, which was based on programs through April 2022.

an investment tax credit and production tax credit. Fossil fuel energy instead faces a carbon tax where τ_t^f is a carbon tax and κ is a technological constant relating electricity generated from fossil fuels to carbon emitted. It is worth noting that a carbon tax is identical to an appropriately scaled negative production tax credit. Revenues raised from a carbon tax are rebated lump sum.

The household's optimal choice for clean energy and fossil fuel investment are now given by two Euler equations:

$$p_t^c\left(1-\tau_t^{inv}\right)\lambda_t = \beta\lambda_{t+1}\left[\begin{array}{c} G_1^c\left(K_{t+1}^c\right)\left(p_{t+1}^e+\tau_{t+1}^p\right) \\ +\, p_{t+1}^c\left(1-\tau_{t+1}^{inv}\right)\left(1-\delta_c\right)\end{array}\right] \text{and} \tag{9}$$

$$p_t^f\lambda_t = \beta\lambda_{t+1}\left[G_1^f\left(K_{t+1}^f\right)\left(p_{t+1}^e+\tau_{t+1}^f\kappa\right)+p_{t+1}^f\left(1-\delta_f\right)\right], \tag{10}$$

where λ_t is the household's marginal utility of consumption. In steady state, the fossil fuel and clean energy capital stocks are jointly determined by conditions equating the user cost of capital to its marginal product, analogous to the case with just clean energy capital:

$$p_c\left(1-\tau_{inv}\right)\left(r+\delta_c\right) = G_1^c\left(K_c\right)\left(p_e+\tau_p\right) \text{ and} \tag{11}$$

$$p_f\left(r+\delta_f\right) = G_1^f\left(K_f\right)\left(p_e+\tau_f\kappa\right). \tag{12}$$

If both generation functions G^f and G^c are constant return to scale, then the marginal increase in electricity production is constant for each unit of capital. In that case, only one of the steady-state conditions can hold in equilibrium (we are implicitly assuming that K_c and K_f are positive in steady state). In effect, a corner solution obtains in the long run, and only clean or fossil fuel technology is utilized. Increases in carbon tax and ITC/PTC would have an impact on this choice only if the unsubsidized (utilization-adjusted) user cost for fossil fuel generation was lower than the user cost and the carbon tax or ITC/PTC is large enough to make the relative utilization-adjusted user cost for clean energy lower. However, to the extent that clean energy has a lower user cost, further increases in the carbon tax or ITC/PTC do not have an impact on the steady-state level of capital stock. The level of capital stock is pinned down by the price of electricity p_e, which falls as the power generation capital stock rises.

If the generation functions G^f and G^c exhibit decreasing returns to scale, then both technologies may be utilized in steady state. In this case, incremental increases in either the carbon tax or the ITC/PTC will shift the energy mix toward clean energy. Nevertheless, the policies are not fully interchangeable so long as we rule out negative carbon taxes or negative subsidies. While either policy could achieve a given target for the mix of clean and fossil fuel generation, the overall level of electricity production rises under an ITC/PTC policy and falls under a carbon tax policy. The emissions reduction is, hence, greater under a carbon tax, but output, consumption, productivity, and wages all fall relative to those in the ITC/PTC policy.

The emissions difference between a clean energy subsidy policy and a carbon tax depends on the price elasticity of electricity demand. Relative to a carbon tax, subsidies encourage electricity consumption and discourage conservation. If household and industrial demand for electricity is sensitive to price, a carbon tax will have a relatively large effect on electricity consumed and hence emissions. By contrast, a subsidy policy—by encouraging electricity consumption—would partially undo the switch from fossil fuel to clean energy by raising overall electricity consumption (Holland, Hughes, and Knittel 2009).

One argument against carbon taxes is that these taxes have an adverse impact on poor households with inelastic energy consumption and whose energy consumption is a larger share of household expenditure. So long as *absolute* energy consumption is increasing in household income (see figure 12), a carbon tax recycled as a lump sum dividend provides poor households sufficient resources to both maintain their pretax energy consumption *and* increase non-energy consumption.

V.B. Optimal Policy

The difference in subsidy versus carbon tax policy in terms of electricity prices and electricity consumption begs the question of whether a subsidy policy is economically preferable to a carbon tax. Is there a case for a clean energy subsidy in lieu of a carbon tax? To address this question, we augment our baseline model to include damages from emissions and consider the planner's problem. As before, we assume only emissions from electricity production, ignoring transportation or land use. The details of the planner's problem are available in online appendix A.

The planner's choice from clean energy and fossil fuel capital are given by

$$p_t^c = \frac{1}{1+r_t}\left[p_{t+1}^e G_c'\left(K_{t+1}^c\right) + p_{t+1}^c\left(1-\delta_c\right)\right]$$

$$p_t^f = \frac{1}{1+r_t}\left[p_{t+1}^e G_f'\left(K_{t+1}^f\right) + p_{t+1}^f\left(1-\delta_f\right)\right] - \underbrace{\mu_{t+1}\kappa G_f'\left(K_{t+1}^f\right)}_{\text{time-varying carbon tax,}}$$

where μ_{t+1} is the multiplier that implicitly prices the damages from cumulative carbon emissions.

It is clear that the optimality condition for clean energy capital is the same as what would obtain with zero subsidies; that is, even in the presence of fossil fuel damages, the planner's allocation for clean energy capital is unchanged relative to the household. The condition that is distorted is the choice of fossil fuel capital, with the planner taking into account that additional fossil fuel capital increases damages from emissions. To implement the planner's allocation, the fiscal authority would need to levy a carbon tax that enters into the Euler equation for fossil fuel capital (the final expression in the equation). Note that we have made assumptions on neither the elasticity of electricity demand nor the relative price of clean versus fossil fuel capital.

Since emissions are cumulative and production exhibits diminishing returns in electricity, the (asymptotic) steady state features zero fossil fuel capital and a carbon tax high enough to ensure that only clean energy capital is utilized. The price of electricity will rise to incentivize increases in clean energy capital investment along with reductions in electricity demand.

Why does the planner rely exclusively on the price of electricity to incentivize a switch to clean energy power generation? The intuition is that subsidies do not change the underlying resource cost of clean energy capital. The incentive to choose power generation via clean versus fossil fuel capital depends only on the relative technological cost (i.e., p_t^e versus p_t^f). The only externality comes from damages generated by reliance on fossil fuel power generation. The main benefit of fossil fuels is power generation, and this benefit must be weighed against damages from emissions; a single instrument is sufficient for correcting that externality.

V.C. Learning by Doing Externality

In the context of our model, introducing a learning-by-doing externality can restore scope for a clean energy subsidy (in addition to a carbon tax). Households now do not internalize the impact of their investment on the price of capital and therefore underinvest relative to a social planner. This can be seen by comparing the household's Euler equation for clean energy investment (in the baseline case where the price of capital is exogenous) with that of the planner who internalizes that faster investment results in a

faster decline in price of capital. Indeed, a subsidy could be warranted even if the unsubsidized price is low enough today or expected to be low enough in the future that clean energy capital is the only power generation source in the long run.

The Euler equation for investment from the planner now differs from the private optimality condition that does not take account of the learning-by-doing externality:

$$p\left(K_t^c\right)u_c\left(C_t\right) = \beta u_c\left(C_{t+1}\right)\left[\begin{array}{c} p_{t+1}^e G_c\left(K_{t+1}^c\right) + p\left(K_{t+1}^c\right)\left(1-\delta_c\right) \\ - p_c\left(K_{t+1}\right)I_{t+1}^c \end{array}\right]. \tag{13}$$

Under learning by doing, there is an additional marginal benefit to an added unit of clean energy capital—a lower price of future investment (the final expression in the equation). The marginal benefits from increasing capital are the discounted sum of the added electricity generated, the market value of the undepreciated capital stock, and the decrease in the cost of new investment. This last term is given by the change in the price of the capital stock multiplied by next period expected investment. This last term does not appear in the household's private Euler equation and justifies a time-varying investment credit to internalize the learning-by-doing externality.[44]

V.D. Efficiency Impacts of Tax Credits versus Carbon Pricing

Our model abstracts from many dimensions of difference between subsidies and carbon pricing. One important difference is that pricing carbon, depending on how it is implemented, could generate revenue for the government. These revenues could be used to offset other distortionary taxes (Barron and others 2018; Goulder 1995), address equity concerns (Goulder and others 2019), or help achieve other policy objectives. A subsidy-based approach costs the government the subsidy amounts and imposes the marginal cost of raising government funds on the economy.

Our model also abstracts from differences in carbon emissions between unsubsidized energy resources. In practice, these can vary considerably, meaning a single clean energy subsidy does not reflect the fact that the benefits of zero-carbon power sources will vary depending on which

44. For a steady state to exist, it must be the case that $p_c(K) = 0$ for sufficiently large levels of capital (i.e., there must be some diminishing returns to learning).

unsubsidized energy resources they displace. For example, hydropower plants generate electricity without emitting CO_2, while coal plants emit about a metric ton of CO_2 per MWh, meaning that at a social cost of carbon of approximately $200 per ton, carbon emissions raise the cost of coal-fired electricity by several multiples. Coal and natural gas generation have different emissions intensities, and even within a fuel type, there is considerable heterogeneity in emissions rates (Kotchen and Mansur 2014). In the transportation sector, emissions are a function of vehicle fuel economy, which also varies considerably. Under the IRA, clean energy that displaces zero-carbon energy such as hydropower is subsidized at the same rate as clean energy that displaces the dirtiest resources.

In principle, this issue could be addressed by adjusting production tax credits based on regional or temporal characteristics that are correlated with the emissions rates of the unsubsidized energy (Abrell, Rausch, and Streitberger 2019). For example, EIA shows that the electricity grid produces emissions about eight times higher in Wyoming than Washington State (McArdle and Fasching 2022), so the PTC could be increased for clean energy producers that locate near Wyoming and reduced for producers in the Pacific Northwest. In practice, it may be difficult to legislate accurate adjustment factors given changing conditions on the electricity grid.[45] Further, it may be politically challenging to reward investments in some politicians' constituencies more than others.

Other provisions of the IRA subsidize the energy-using or energy-producing asset, irrespective of how much it is operated. The investment tax credit for zero-carbon electricity subsidizes the construction of the facility rather than its operation. Relative to the PTC, the ITC provides a lower incentive to produce clean energy once the facility is constructed and thus a lower incentive to locate in areas with the highest production potential (Aldy, Gerarden, and Sweeney 2018). With lower capital costs for relatively mature technologies like wind and solar that are expected to be deployed with the IRA, many developers could opt for the production tax credit, which could minimize such distortions. Similarly, the electric vehicle tax credits subsidize vehicle purchases without regard to how much electric vehicles are driven. Electric vehicles that are used as second cars and driven less will offset fewer emissions than vehicles that replace a household's only car.[46]

45. Note that efficiency gains from differentiated subsidies across technologies may be limited in practice (Abrell, Rausch, and Streitberger 2019).

46. Burlig and others (2021) use electric vehicle charging data to show that vehicles in California through 2019 were driven substantially less than vehicles with internal combustion engines.

Overall, a shortcoming of fixed tax credit rates for supply- and demand-side resources is that they are relatively inflexible as technology and market conditions change (Peñasco, Anadón, and Verdolini 2021).[47] Carbon pricing enables households and businesses to select their preferred approaches to lower emissions, which can help to reduce costs and account for other welfare-relevant considerations that vary across individuals and firms. Carbon pricing also can enable coordination across sectors and geographies. When policy stringencies differ across sources and locations, emissions leakage can occur, though there are several policy options to mitigate leakage when policies are not harmonized (Böhringer and others 2022).

Another potential rationale for policy instruments that lower electricity prices could be that retail prices exceed social marginal costs. Borenstein and Bushnell (2022) argue that residential electricity rates are higher than the full social marginal costs in many locations across the United States, including the Southwest and Northeast. However, their estimates for external marginal costs are based on older values of the social cost of carbon and marginal damages of criteria pollutants, which are considerably lower than more recent estimates (Rennert and others 2022; Shindell and others 2021). In addition, federal tax credits are relatively blunt instruments to correct for retail pricing distortions that dominate only in a few regions, especially since tax credit uptake and retail rate impacts are not necessarily correlated with areas with retail price distortions. Also, the analysis in Borenstein and Bushnell (2022) looks only at residential electricity rates, and marginal rates for other sectors are generally lower.

In addition to the negative externalities from emissions, climate change is also associated with positive innovation-related externalities, particularly since carbon emissions are unpriced in many parts of the world (van den Bergh and Savin 2021; Gillingham and Stock 2018; Acemoglu and others 2012; Popp, Newell, and Jaffe 2010). Both carbon pricing and subsidies are aimed at addressing not only negative externalities from emissions but also positive innovation-related externalities. Induced innovation, economies of scale, network effects, and learning-by-doing effects can be altered by subsidies, carbon pricing policies, and other instruments, which can lower costs of low-emissions technologies and of energy services more broadly, though impacts depend on policy design. For example, subsidies for nascent technologies, like electric vehicles, may push producers down a learning

47. On the other hand, features in the IRA such as the qualifying emissions threshold for the power sector PTC and ITC illustrate how dynamic elements could be incorporated into subsidy design.

curve. Or, with more electric vehicles on the road, the economics of installing and operating charging stations improve, and the number of mechanics with expertise working with electric vehicles will increase. Future research should elaborate on the advantages of subsidies versus those of carbon pricing for incentivizing innovation. It is plausible, for example, that subsidies involve less uncertainty for investors in clean technology and are therefore better at addressing liquidity constraints.

V.E. Estimates Comparing Tax Credits with Carbon Pricing

Several modelers have simulated the reductions achieved with a carbon price and compared them to tax credits. For example, Roy, Burtraw, and Rennert (2021) find that even relatively modest carbon fees reduce emissions more than the types of tax credits that were included in the IRA and that the two policies together can achieve greater emissions reductions at a lower fiscal cost, while also insulating households from increased costs.

Here, we use the energy-economic model US-REGEN to investigate how electric sector outcomes vary between a scenario with IRA incentives and another that matches annual electricity CO_2 emissions without the IRA (implicitly assuming a cap-and-trade policy approach). Table 3 compares electricity generation shares by technologies across these scenarios, along with emissions, electricity prices, and abatement costs. This comparison illustrates how carbon pricing leads to lower coal generation relative to a subsidy-focused approach, since the latter does not distinguish between the carbon intensity of unsubsidized generation, which has implications for associated air quality co-benefits. These lead to 68 percent reductions in power sector CO_2 emissions from 2005 levels by 2035. Average abatement costs in the CO_2 equivalent policy are relatively low ($\$10/tCO_2$), given the low incremental costs of coal-to-gas switching and renewables deployment at these levels.[48] Table 3 also confirms that tax credits lead to lower electricity prices, relative to carbon pricing with equivalent CO_2 emissions. These lower prices are due to the prevalence of subsidized resources that put downward pressure on electricity prices and shift costs from ratepayers to the federal government (i.e., taxpayers).

While these estimates pit the subsidies against an idealized alternative policy, another useful comparison is between the IRA subsidies and the social cost of carbon, which essentially measures whether the subsidies

48. Average abatement costs are presented for comparability between the IRA and carbon pricing scenarios. Marginal abatement costs under the carbon pricing scenario are $\$12–15/tCO_2$ between 2030 and 2035, which are the shadow prices on the emissions cap constraint.

Table 3. Comparison of Electric Sector Metrics in the IRA Scenario and Carbon Tax Scenario That Matches the IRA CO_2 Emissions Time Path

		IRA scenario		*Carbon tax*		*Difference (percentage points)*	
Metric (units)	*2021*	*2030*	*2035*	*2030*	*2035*	*2030*	*2035*
Generation share (%)							
Coal	22	11	8	7	4	–4	–5
Coal CCS	0	3	3	0	0	–3	–3
Gas	39	20	18	35	34	15	17
Gas CCS	0	0	0	0	0	0	0
Other	2	9	11	7	8	–2	–3
Nuclear	19	17	14	17	16	0	2
Hydro	6	6	6	6	6	0	0
Wind and solar	13	33	41	28	32	–6	–9
CO_2 (% drop from 2005)	35	64	68	64	68	0	0
Generation price ($/MWh)	$64	$56	$52	$65	$62	16%	20%
Abatement cost ($/t$CO_2$)	n/a	$45–61	$45–61	$10	$10	–85%	–82%

Sources: US Energy Information Administration; Environmental Protection Agency; and authors' calculations.

Note: Historical electricity generation shares come from the US Energy Information Administration, "Electric Power Monthly," https://www.eia.gov/electricity/monthly/. Historical emissions come from United States Environmental Protection Agency, "Inventory of U.S. Greenhouse Gas Emissions and Sinks," https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks.

pass a cost-benefit test. Several analyses estimate the cost-effectiveness of components of the IRA, permitting this comparison. Analyzing a suite of tax credits for the power sector like the ones included in the IRA (the credits that were included in the Build Back Better Act passed by the House in fall 2021), Greenstone and others (2022) estimate that the credits would reduce emissions at a cost of \$33–50/tCO_2, substantially below the most recent estimates of the social cost of carbon of approximately \$200/tCO_2 (EPA 2022). Similarly, Stock and Stuart (2021) compare a suite of electric sector policies to the social cost of carbon and find that extensions of the PTC, ITC, and subsidies for CCS (but not all of the credits eventually included in the IRA) have an average abatement cost of about \$35/tCO_2, well below the social cost of carbon. Cole and others (2023) estimate that the electric vehicle tax credits, combined with the subsidies for charging stations included in the Infrastructure Investment and Jobs Act, would reduce emissions at a cost of approximately \$95/tCO_2.

As shown in table 3, our estimates in US-REGEN indicate that IRA tax credits reduce CO_2 emissions at an average abatement cost of \$45–61/tCO_2 for the power sector (discussed in section II.D). There may be an efficiency gap between IRA incentives and carbon pricing with equivalent CO_2 (with

average abatement costs of about $\$10/tCO_2$), but these incentives nevertheless pass the benefit-cost test for most of the updated ranges for the social cost of carbon (Rennert and others 2022), even before accounting for air pollution and other potential co-benefits.

V.F. Political Economy Considerations and Alternative Policy Instruments

The approach embodied in the IRA is motivated in part by political economy constraints on feasible policy instruments in the 117th US Congress, including legislative dynamics that led to climate policy via budget reconciliation, which is a procedure to pass budgetary legislation that can override filibuster rules in the Senate and hence can pass by a simple majority rather than a sixty-vote supermajority.

Although carbon pricing approaches can be efficient, effective, and equitable, their strengths can create political liabilities by raising costs of energy. Many Americans support government action to address climate change, but willingness to pay may be low (Jenkins 2014).[49] In contrast, tax credits can lower energy prices and hide policy costs, which may be one reason why subsidies tend to poll better in the United States, relative to carbon pricing (Krosnick and MacInnis 2020; Bergquist, Mildenberger, and Stokes 2020).

In addition to tax credits and carbon pricing, there are several additional policy instruments that have been used and proposed to reduce emissions and encourage adoption of clean energy technologies, including rate-based performance standards (e.g., the Clean Power Plan proposed by the US EPA in June 2014), portfolio standards (e.g., for renewables or broader clean energy), mandates (e.g., for offshore wind and energy storage in various US states), feed-in tariffs (e.g., for renewables in several European countries), and others. Each instrument has policy design elements such as stringency, timing, trading provisions, and eligible technologies that affect economic and environmental outcomes and that emphasize different abatement margins. The literature suggests that the relative performance of tax credits versus that of other policy instruments depends on several factors about the setting, including the regional energy system, level of decarbonization, renewable resource quality, and demand effects (Borenstein and

49. For instance, a poll by the AP-NORC Center and EPIC in 2019 indicated that 43 percent of US adults are unwilling to pay an additional $1 on their monthly utility bill; Energy Policy Institute at the University of Chicago, "Infographic: Where Americans Stand on Energy and Climate," https://epic.uchicago.edu/news/infographic-where-americans-stand-on-energy-climate/?_paged=4.

Kellogg 2022; Abrell, Rausch, and Streitberger 2019; Young and Bistline 2018; Paul, Palmer, and Woerman 2015; Fell and Linn 2013).

V.G. Industrial Policy Components

As discussed in section I.B, some of the tax credits include bonuses for using domestic content or are available only for domestically produced goods with stringent sourcing requirements, as is the case for the electric vehicle credits for purchases (not leases). There are also significant bonuses for certain labor practices. While the political economy benefits of the bonuses are clear, the economic implications depend on underlying conditions. One central question is how much the bonuses will lead to adjustments. For example, if most electricity plant construction workers and operators already receive prevailing wages and use apprentices, the bonuses serve as a political statement but will not meaningfully change practices or the economic costs. If the bonuses lead to behavioral adjustments, they may be solving market failures. For example, the literature on industrial policy suggests governments can use temporary protection to help local industries achieve economies of scale (Juhász 2018). Also, supply chain vulnerabilities may create externalities if individual buyers do not fully account for the broader economic harm created by disruptions in foreign supply. In the absence of market failures, the bonuses could raise costs more than is socially beneficial, undermining the climate benefits of the tax credits. Future work quantifying these possible externalities will be valuable.

Domestic content provisions also have implications for trade partners, as evidenced by the reaction to the IRA from European leaders. Some of the United States' most important trade partners have sizable carbon prices (e.g., the European Union and Canada), and those countries may feel pressure from industry to reduce those costs, lest they lose production to the United States. Such a scenario would undermine the climate benefits of the IRA.[50] Clausing and Wolfram (2023) discuss the possible dynamics when countries adopt asymmetric approaches to carbon mitigation.

VI. Quantifying the IRA's Macroeconomic Impacts

VI.A. Recent Trends in Electric Power Investment

As shown in section II (figure 4), the IRA has significant impacts on the level of investment in clean electricity generation, with 34–116 gigawatts

50. If the European Union or other countries respond by introducing clean energy subsidies rather than reducing carbon taxes, global climate benefits could be strengthened.

Table 4. Investment in Power Generation Structures and Equipment

	Nominal, 2018–2022 averages			*Growth, annualized rate* 2012–2022		*US-REGEN IRA impact* *Ten-year average*
	$ billion	*% of BFI[a]*	*% of GDP*	*Real output*	*Price level*	*$ billion (2022)*
Gross domestic product	22,350			2.2	2.5	
Nonresidential fixed investment	2,974		13.3	3.3	1.5	
Structures	633	21.3	2.8	–0.8	4.2	
Electric power structures (BEA estimate)	79	2.7	0.4	–3.1	4.2	21
Equipment	1,199	40.3	5.4	2.5	0.6	
Electrical transmission, distribution, and industrial apparatus	52	1.8	0.2	3.8	2.4	7
Electrical equipment, not elsewhere classified	9	0.3	0.0	1.2	2.1	

Sources: Bureau of Economic Analysis and authors' calculations.
a. Business fixed investment

of nameplate capacity added annually on average relative to 18 GW/year on average in the previous decade and 36 GW/year in 2021. Table 4 shows data from the Bureau of Economic Analysis (BEA) for investment in power generation and electricity distribution. Over the last five years, nominal investment in electric power structures averaged $79 billion, and nominal investment in electrical transmission and distribution averaged $52 billion. Investment in electric power structures accounts for about one-eighth of overall structures investment, but nonresidential structures investment is quite modest as a share of GDP. Likewise, electrical transmission and distribution account for about 4 percent of nonresidential investment in equipment. Total investment in electrical power generation, distribution, and transmission is about 5 percent of nonresidential fixed investment and less than 1 percent of GDP.

The US-REGEN model sees a boost (relative to 2022 levels) of approximately $21 billion per year over ten years in electric power generation and approximately $7 billion per year in transmission and distribution. These magnitudes are sizable relative to the current level of investment, but comparatively small as a share of both overall investment and overall economic activity. Even substantially larger increases in investment in power generation, transmission, and distribution (a doubling or more) would carry

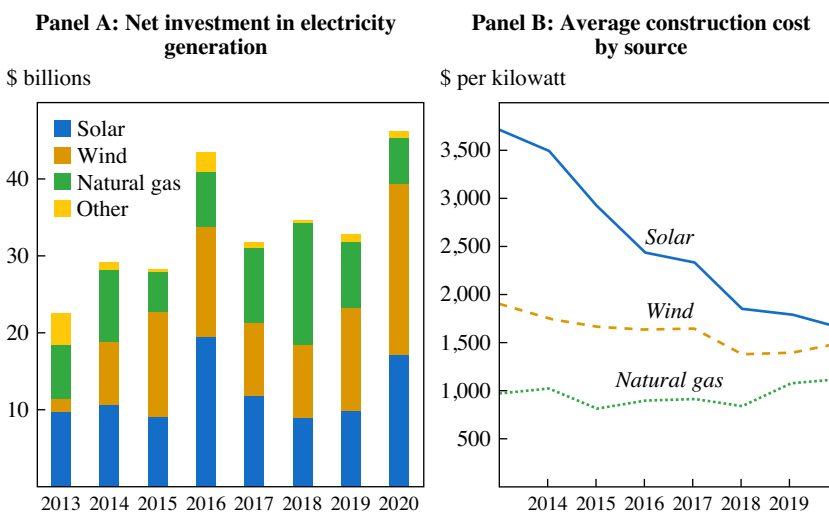
relatively modest macroeconomic impacts given the low share of power and electricity investment relative to overall investment. An apt comparison for the magnitude of the IRA impacts on electric power generation could be the shale oil revolution in the prior decade. Nominal investment in mining and wells rose from $98 billion between 2006 and 2010 to $152 billion between 2011 and 2015—an increase of more than 50 percent in a short period of time. Investment in mining and oil field machinery nearly doubled from $18 billion to $33 billion. The macroeconomic impacts from this investment boom appeared comparatively modest, however, with relatively limited macroeconomic impacts from the sharp slowdown in shale oil investment after 2015.[51]

The BEA data also suggest relatively high price growth over the past decade for investment in electricity generation. As table 4 shows, prices rose 4.2 percent (annualized rate) for investment in electric power generation structures, and real investment fell 3.1 percent (annualized) over the same period. Price increases were more in line with overall inflation for transmission and distribution, rising 2.4 percent at an annualized rate. In contrast to real investment in electric power structures, real investment in transmission and distribution rose over the past decade. Price increases for investment in structures have been particularly sharp over the last two years; the price index for electric power structures rose 12.3 percent year-over-year as of 2022:Q4. Negative real investment in electric power generation over this period is largely a function of flat nominal investment and surging price indexes.

The BEA investment data are mirrored by data collected by the US Energy Information Administration (EIA) on net additions to electricity generation and data on construction cost of new electric generation facilities. Figure 16 shows total construction expenditures for solar, wind, and natural gas electricity generation from 2013 to 2020. Total investment in power generation rose from $22 billion to $46 billion, with solar and wind rising from $10 billion and $2 billion to $17 billion and $22 billion, respectively, over this period. Nameplate capacity installed rose from 12 GW to 31 GW during this time—much faster than the level of real investment inferred from BEA data. Figure 16 also shows substantial drops in the

51. Bureau of Economic Analysis, "National Data: National Income and Product Accounts," tables 5.4.5 and 5.5.5; https://apps.bea.gov/iTable/?reqid=19&step=2&isuri=1&categories=survey. US GDP growth did decelerate in 2015, and nonresidential structures investment turned sharply negative (falling nearly 10 percent year-over-year). But the unemployment rate continued to fall, and core inflation appeared largely unchanged.

Figure 16. Trends in Electricity Generation Investment and Costs



Source: US Energy Information Administration.

average construction cost of solar power over this period, along with substantial declines for wind generation. The construction cost for natural gas remained largely stable over this period.

How do we reconcile the BEA and EIA data? The BEA reports gross investment in structures and equipment. The BEA's annual data on net investment from its fixed asset tables are quite close to the EIA's value for net investment in electricity power generation. In 2020, the BEA recorded gross investment of $83 billion in electric power structures and $41 billion in depreciation; the EIA's value for net investment in 2020 was $46 billion.[52] The EIA, however, does suggest materially different trends in both real investment and the price index for investment in electric power generation due to a sharp fall in construction costs for utility scale solar and wind.[53]

52. Bureau of Economic Analysis, "National Data: Fixed Assets Accounts Tables," tables 2.5 and 2.7; https://apps.bea.gov/iTable/?reqid=10&step=2&isuri=1.

53. The difference in price trends in BEA and EIA data may reflect that the BEA price index is weighted on a capital stock basis (i.e., disproportionately weighted toward coal and gas). The BEA uses indexes from Handy-Whitman and the Bureau of Reclamation to construct its price index for electric power structures. It is not clear that those price indexes have sufficient weight in solar and wind, which account for a sizable share of electric power investment in the recent data. The EIA documents an average 3.1 percent increase in construction costs for natural gas power plants, which is close to the 2.6 percent increase in the BEA's price index.

The US-REGEN model shows substantial and growing impacts on electricity prices relative to the non-IRA scenario (figure 11 in section III). To a first approximation, the inflation impact of declining retail electricity prices is simply its weight in household prices multiplied by the change in electricity prices (post-IRA). The US-REGEN model finds that retail electricity prices are 1.2 percent higher in 2025 (relative to baseline) but fall 2.2 percent by 2030 and 12.8 percent by 2050. Electricity has a weight of 2.5 percent in the Consumer Price Index (CPI) and 1.3 percent in the price index for personal consumption expenditures. Therefore, electricity prices would add 1.5–3 basis points to inflation in 2025 but subtract 3–6 basis points in 2030 and 15–30 basis points in 2050. Overall, these are small direct effects on inflation. For reference, the impacts on the price of electricity in 2025 and 2030 are an order of magnitude lower than the increase in retail electricity prices experienced over the pandemic.[54]

VI.B. FRB/US Simulation

To quantify the macroeconomic impacts of the climate provisions of the IRA, we rely primarily on the Federal Reserve's US model (FRB/US). This general equilibrium model is regularly estimated and used by Federal Reserve staff in formulating forecasts and assessing the macroeconomic outlook. Ideally, one would be able to model economic and energy market impacts within a single general equilibrium model, where changes in subsidies or a carbon tax would both have a joint impact on industry equilibrium and incorporate feedbacks to the broader economy. The current FRB/US model has only limited modeling of energy market impact on the broader economy (primarily through the price of oil). To simulate the macroeconomic impact of the IRA's climate provisions, we take the principal economic outputs of the US-REGEN model and incorporate those impacts into the current baseline FRB/US model.

Specifically, the tax credits received by households for electric vehicles and for residential improvements (heat pumps, etc.) are modeled as an increase in transfer income to households (akin to cash stimulus), and the increased investment in wind, solar, and other clean power generation is modeled as an increase in the growth rate of business fixed investment.[55] Additionally, we also include a shift in the CPI for energy to reflect the

54. From 2020 to 2022, the CPI for electricity services has risen 22.6 percent.

55. To the extent that household rebates are captured by manufacturers through higher prices, the transfer raises corporate profits and equity valuations. The marginal propensity to consume out of the transfer would likely be lower.

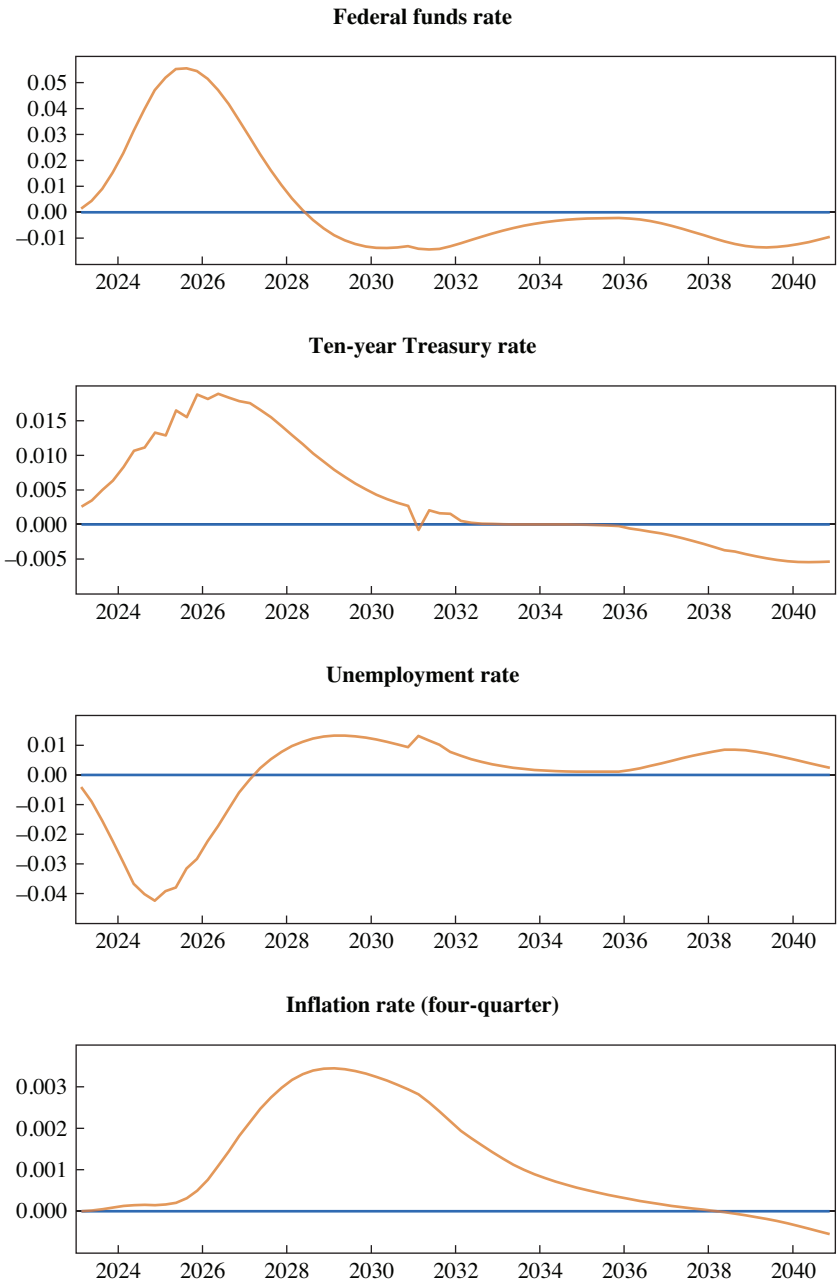
impact of lower retail electricity prices from increased electricity production. The outputs from the US-REGEN model are at five-year time steps, so we convert these values to quarterly shocks for the FRB/US model. The impulse responses below show the relative impact of the IRA on the federal funds rate, the unemployment rate, ten-year Treasury rates, and the inflation rate.

The FRB/US simulation largely confirms what might be expected given the small share of investment accounted for by power generation and the relatively modest size of the household transfers. As figure 17 shows, the federal funds rate rises initially due to stronger nonresidential investment and increased household consumption from clean vehicle and residential improvement tax credits. The federal funds rate and ten-year Treasury rate peak by 2026 and then return to their baseline levels. The funds rate falls slightly below baseline after 2030 as the fiscal impetus from increased nonresidential investment and increasing transfers turns to a drag after 2030. The unemployment rate initially falls before rising above baseline slightly after 2030. Quantitatively, all effects are small. At its peak, the federal funds rate increases about 6 basis points, and the Treasury rate increases about 2 basis points. The maximum fall in the unemployment rate is approximately 4 basis points. Impacts on core inflation are an order of magnitude smaller in the simulation, with core inflation rising relative to baseline over the first decade before falling after 2030. However, headline inflation falls because of the direct effect of lower electricity prices on consumer prices, which the FRB/US model does not include. Including these direct effects on electricity, headline CPI inflation falls about 3–6 basis points by 2030 and up to 30 basis points by 2050.[56]

The FRB/US model may understate the impacts from higher nonresidential investment in power generation in two ways: (1) the model may not fully capture the upstream impacts on manufacturing and materials demand from an increased level of structures and equipment investment; (2) the model is unlikely to capture fluctuations in energy commodity prices (electricity, natural gas, crude oil, and gasoline) that will have material impacts on producer and consumer prices. Using data from BEA input-output tables, we see that the upstream impacts of investment in power structures do not appear significantly larger than demand for any other commodity. The 2012 total requirements table shows that a $1 increase in final demand for investment in power structures implies a $1.64 increase

56. Del Negro, di Giovanni, and Dogra (2023) examine how differential price stickiness and subsidies may lead a green energy transition to be disinflationary.

Figure 17. Interest Rate, Unemployment, and Inflation Rate Response to IRA



Source: Authors' calculations using FRB/US.
Note: Macroeconomic effects of funds rate perturbation (VAR expectations) shown.

in gross output across all commodities.[57] Upstream impacts for electrical transmission and distribution equipment are somewhat higher: $2.46 for each $1 increase in final demand. By comparison, the ratio of gross output to GDP is approximately 1.8.

As we noted earlier, the scale of the increase in fixed investment under the IRA is comparable to the shale oil boom in the early 2010s. The direct contribution of shale oil investment in mining structures and equipment may not have been large (relative to aggregate investment or GDP), but it clearly had significant effects on energy prices that kept overall inflation low and supported a recovery in manufacturing after the financial crisis. Moreover, the sharp drop in global oil prices, slowdown in shale oil investment after 2014, and resulting weakness in US manufacturing stalled the exit from the zero lower bound until 2016.

The macroeconomic impact of the clean energy provisions of the IRA should also be considered in concert with the other major recent legislation that increases demand in domestic manufacturing and construction. Both the Infrastructure Investment and Jobs Act and the CHIPS and Science Act have provisions that are intended to increase structures and equipment investment, with expenditures ramping up on a similar timeline to that of the clean energy investments due to the IRA. The combined impact of this increased investment demand, along with upstream impacts on manufacturing, construction, and raw materials, may carry a more meaningful quantitative impact on the US macroeconomic aggregates. One data point that points to a larger macroeconomic impact is announcements for battery manufacturing since the passage of the IRA; a tabulation finds twenty-nine companies announcing almost $50 billion in domestic battery manufacturing as of March 2023. These investments in battery production that may be eligible for IRA manufacturing tax credits have not been incorporated in our quantitative estimates.

It's important to note that any negative macroeconomic effects on fossil fuel extraction, refining, and utilities are not modeled here, nor are any broader macroeconomic effects for the revenue components of the IRA. Power investment in structures and equipment is small relative to overall economic activity, and so are fossil fuel extraction and refining. Moreover, oil and natural gas are partly global commodities whose outlook

57. Business fixed investment in FRB/US is not modeled at the industry level, so upstream impacts to manufacturing are captured only indirectly, through lagged terms and accelerator effects. Any impact of lower electricity prices on manufacturing is also not captured.

will be strongly influenced by global events (e.g., European demand for liquified natural gas), with demand for natural gas in the United States likely to remain steady in any case. Employment in these industries is small, and investment in fossil fuel power generation (particularly coal) was already waning well before the IRA.

VII. Impacts of the Macroeconomic Environment on the IRA

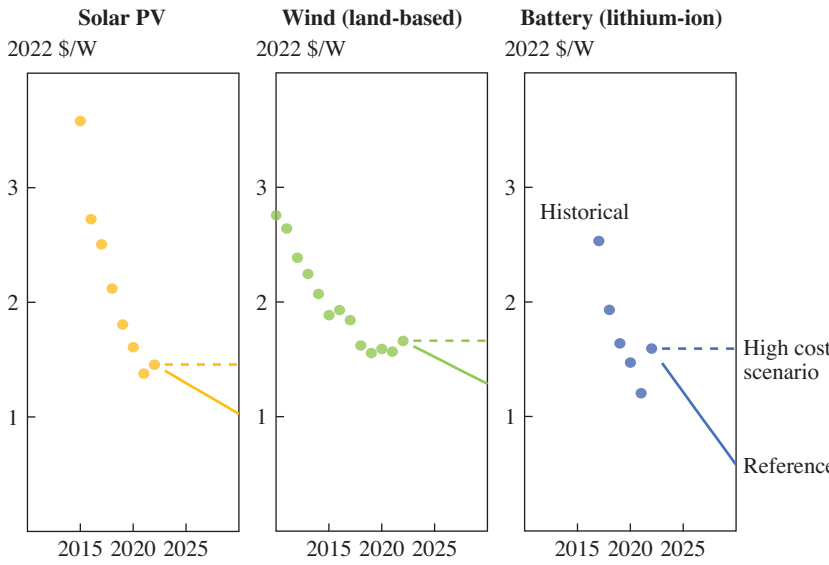
To test the sensitivity of IRA impacts to the assumed macroeconomic environment, we compare the US-REGEN electric sector outputs from earlier sections to a scenario with higher supply-side costs and interest rates. In this illustrative scenario, the discount rate for power sector investments is increased from 7 percent in the reference run to 11 percent (Dunkle Werner and Jarvis 2022; EPA 2018). In addition, this higher cost scenario assumes that upward pressure on labor and materials costs leads to increases in the capital costs of generation and energy storage technologies. Elevated costs in 2022 are assumed to persist through 2030 instead of following the declining cost trajectories used in earlier sections (figure 18), as discussed in Bistline, Roney and others (2023).

This stylized scenario with a more pessimistic macroeconomic outlook and higher costs leads to higher emissions and lower clean electricity generation relative to the scenario with reference costs presented earlier (figure 19). Higher interest rates increase the costs of new investments, especially for capital-intensive technologies such as many IRA-subsidized, zero-emitting resources (as the LCOE examples in section III illustrate). This dynamic lowers the deployment of wind, solar, and CCS-equipped capacity and increases generation from existing assets, especially coal- and gas-fired capacity.[58] These changes lead to increases in CO_2 emissions in the higher cost scenario, which are 1,190 $MtCO_2$/year in 2030 (compared with 870 $MtCO_2$/year with reference costs).

Table 2 in section II compares emissions and fiscal costs associated with this scenario with those of the central IRA case. Economy-wide emissions reductions are similar to reference levels with higher costs, and cumulative tax credit expenditures through 2031 are $244 billion (i.e., similar to the CBO/JCT score) instead of $781 billion in the central case.

58. The ratio of wind to solar generation increases with higher costs in 2030, given the relative magnitudes of their capital cost increases (figure 18).

Figure 18. Capital Cost Assumptions for Utility-Scale Solar PV, Land-Based Wind, and Lithium-Ion Battery Storage with Four-Hour Duration



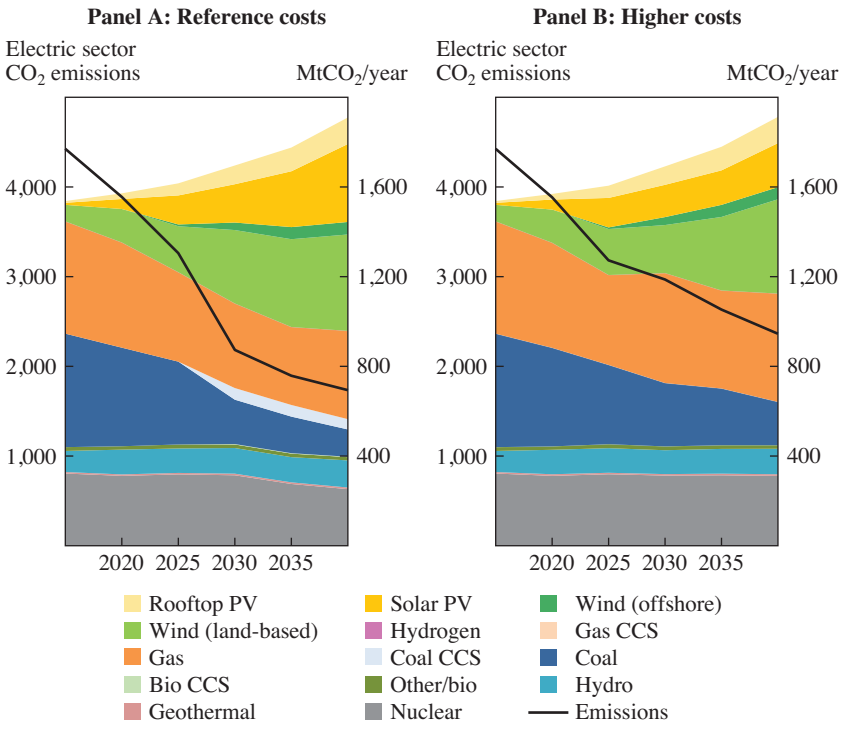
Sources: Lawrence Berkeley National Laboratory; Electric Power Research Institute; and authors' calculations.

Note: Costs are expressed in 2022 dollar terms before accounting for IRA subsidies. Historical costs (dots) for wind come from Wiser and others (2022); for solar, Bolinger and others (2022); and for batteries, Hostert and others (2022). Values for 2022 are based on EPRI (2022) cost estimates.

These comparisons illustrate how macroeconomic conditions may have larger impacts on IRA investments than IRA investments have on macroeconomic conditions, at least for the magnitudes investigated here. There is considerable uncertainty about the persistence of these shocks and their magnitudes, which depend not only on domestic conditions but also on global factors. For instance, prices of materials—including specialty metals and bulk commodities—depend on global material production and demand, which are driven by the pace of decarbonization-related deployment and non-energy demand (International Energy Agency 2023; Wang and others 2023).[59]

59. Another global driver is increasing liquified natural gas exports from the United States, which connect domestic gas markets with global ones. This fuel market integration could lead to a prolonged period of higher US natural gas prices. The scenarios above do not include higher natural gas price sensitivities, which could increase short-run coal generation (and associated emissions) but decrease fossil fuel consumption (and emissions) in the long run (Stock and Stuart 2021; Bistline and Young 2022).

Figure 19. Generation by Technology over Time for a Scenario with Reference Cost Assumptions and with Higher Costs from the US-REGEN Model



Source: Authors' calculations.

VIII. Conclusions

IRA incentives are expected to significantly alter the economics of decarbonization, encouraging deployment of clean energy technologies and lowering emissions. Economic tools will play important roles in the years to come in understanding potential macroeconomic and microeconomic implications of IRA incentives. This paper offers several initial perspectives on what the IRA's climate-related provisions could imply for energy transitions and key macroeconomic indicators, using stylized examples, data analysis, detailed energy systems modeling, and general equilibrium modeling of the economy.

This analysis using EPRI's US-REGEN model suggests that the IRA, along with other policies and market trends, shifts baseline expectations

of firms, households, and policymakers for the pace and extent of future decarbonization.

Clean electricity investments span 34–116 gigawatts of nameplate capacity added annually under the IRA through 2035, compared with 18 GW/year on average in the previous decade and 36 GW/year in 2021 (figure 4 in section II). Average annual additions of low-emitting capacity in US-REGEN are 51 GW/year with the IRA through 2035, which nearly doubles the 27 GW/year in the counterfactual without the IRA.

Electric vehicle sales increase from over 6 percent of new light-duty vehicle sales in 2022 to nearly half of new sales by 2030 (figure 3 in section II). The IRA increases the electric vehicle share of new vehicle sales by 12 percentage points in 2030 from 32 percent in the reference without the IRA to 44 percent with IRA credits.

The projected pace and extent of these changes depend on assumptions about future policies, technologies, and markets. The uncertainty associated with these projections reflects IRA implementation details and unknown responses to siting and permitting challenges, workforce changes, global supply chain shifts, and non-cost barriers to deployment. The IRA continues to drive emissions reductions in this modeling beyond 2030, as key power sector tax credits may last until electricity CO_2 emissions reach 25 percent of their 2022 levels, potentially providing support throughout the 2030s.

The acceleration in the deployment of clean supply- and demand-side technologies in this modeling leads to greater uptake of IRA incentives than initial estimates indicated. These projections indicate that fiscal costs of IRA tax credits for clean electricity, carbon capture, and electric vehicles may be $781 billion by 2031 in our central case—nearly three times the CBO/JCT score for comparable credits. This finding that budgetary effects of the IRA may be larger than initial CBO estimates is reflected in other studies that indicate that US-REGEN estimates in this paper could be conservative for several uncapped tax credits. Notably, the CBO/JCT fiscal score does not reference the implied carbon reductions. Models that indicate emissions reductions in the range of 40 percent imply larger fiscal costs than CBO/JCT. US-REGEN estimates also suggest that the increasing economic competitiveness of transport electrification and renewables in the power sector means that nontrivial shares of these tax credits would be inframarginal transfers.

This analysis points to limited macroeconomic effects of the IRA, though the macroeconomic environment is shown to influence IRA-incentivized

investments. The conceptual framework in section IV provides intuition for fiscal and macroeconomic impacts of the IRA, including the dependence on supply elasticities for labor and key materials, price elasticity of demand for fuels, market and nonmarket bottlenecks, stock dynamics for emissions-intensive assets, and potential for endogenous technical change for low-carbon energy technologies (e.g., learning-by-doing effects). Although changes in investment from the IRA are large in absolute terms (i.e., tens of billions of dollars per year), even substantial investment increases for power generation, transmission, and distribution carry relatively modest macroeconomic impacts given their low shares relative to overall investment. Magnitudes of change in fixed investment under the IRA are comparable to the shale oil and gas revolution in the 2010s.[60] Likewise, household transfers through tax credits for electric vehicles, heat pumps, and so on are relatively modest in size.

Numerical simulations in section VI using the Federal Reserve's FRB/US model with inputs from US-REGEN show the relative impact of the IRA on the federal funds rate, ten-year Treasury rates, and inflation, which rise initially from increases in nonresidential investment and household consumption before returning to their baseline levels. Quantitative effects are small in all cases. Comparisons in section VII using US-REGEN illustrate how increases in the cost of capital and supply-side costs for electricity generation and storage technologies lead to lower IRA-induced investments in low-emitting capacity and higher emissions over time.

Our survey of potential IRA impacts points to several areas for additional research. Ex ante and ex post analysis of individual IRA provisions will be important for updating baselines for future policies and for understanding the effectiveness of IRA incentives. Assessing interactions between IRA incentives and changes in federal regulations, state policies, and company targets will also be important. Future work should also quantify aggregate macroeconomic impacts of the IRA, the Infrastructure Investment and Jobs Act, and the CHIPS and Science Act, as all three are expected to increase investments across a similar timeframe and have impacts on manufacturing, construction, and raw materials. Finally, understanding the economic incidence of subsidies and distributional implications of the IRA

60. Although the direct contribution of shale oil and gas investment may not have been large (relative to aggregate investment or GDP), it clearly had significant effects on energy prices that kept overall inflation low and supported a recovery in manufacturing after the financial crisis.

will be valuable to policymakers and other stakeholders, especially since many IRA provisions target energy equity, environmental justice, and disadvantaged communities.

ACKNOWLEDGMENTS We are grateful to Curtis Carlson, Jason Furman, Ken Gillingham, Arik Levinson, and Jim Stock for helpful suggestions. The authors thank Justin Cooper and Nichole Shepard for research assistance. All remaining errors are our own.

References

Abrell, Jan, Sebastian Rausch, and Clemens Streitberger. 2019. "The Economics of Renewable Energy Support." *Journal of Public Economics* 176:94–117. https://doi.org/10.1016/j.jpubeco.2019.06.002.

Acemoglu, Daron, Philippe Aghion, Leonardo Bursztyn, and David Hemous. 2012. "The Environment and Directed Technical Change." *American Economic Review* 102, no. 1: 131–66. https://doi.org/10.1257/aer.102.1.131.

Aldy, Joseph E., Todd D. Gerarden, and Richard L. Sweeney. 2018. "Investment versus Output Subsidies: Implications of Alternative Incentives for Wind Energy." Working Paper 24378. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24378.

Attwood, J. 2022. "Carbon Capture Could Get $100B in Credits from US Climate Bill." BloombergNEF Insight 29567. https://www.bnef.com/insights/29567/view.

Barron, Alexander R., Allen A. Fawcett, Marc A. C. Hafstead, James R. McFarland, and Adele C. Morris. 2018. "Policy Insights from the EMF 32 Study on U.S. Carbon Tax Scenarios." *Climate Change Economics* 9, no. 1: 1840003. https://doi.org/10.1142/S2010007818400031.

Bergquist, Parrish, Matto Mildenberger, and Leah C. Stokes. 2020. "Combining Climate, Economic, and Social Policy Builds Public Support for Climate Action in the US." *Environmental Research Letters* 15, no. 5: 054019. https://doi.org/10.1088/1748-9326/ab81c1.

Bistline, John E. T. 2017. "Economic and Technical Challenges of Flexible Operations under Large-Scale Variable Renewable Deployment." *Energy Economics* 64:363–72. https://doi.org/10.1016/j.eneco.2017.04.012.

Bistline, John E. T. 2021. "The Importance of Temporal Resolution in Modeling Deep Decarbonization of the Electric Power Sector." *Environmental Research Letters* 16, no. 8: 084005. https://doi.org/10.1088/1748-9326/ac10df.

Bistline, John E. T. 2023. "Power Plant Performance Standards and Tax Credit Interactions: Impacts of Design Decisions and the Inflation Reduction Act on the U.S. Power Sector." White Paper 3002026291. Palo Alto, Calif.: Electric Power Research Institute. https://www.epri.com/research/programs/109396/results/3002026291.

Bistline, John E. T., Robin Bedilion, Naga Srujana Goteti, and Neil Kern. 2022. "Implications of Variations in Renewable Cost Projections for Electric Sector Decarbonization in the United States." *iScience* 25, no. 6: 104392. https://doi.org/10.1016/j.isci.2022.104392.

Bistline, John E. T., Geoffrey J. Blanford, Maxwell Brown, Dallas Burtraw, Maya Domeshek, Jamil Farbes, and others. 2023. "Emissions and Energy Impacts of the Inflation Reduction Act." *Science* 380, no. 6652: 1324–27. https://www.science.org/doi/abs/10.1126/science.adg3781.

Bistline, John E. T., Geoffrey J. Blanford, Trieu Mai, and James H. Merrick. 2021. "Modeling Variable Renewable Energy and Storage in the Power Sector." *Energy Policy* 156:112424. https://doi.org/10.1016/j.enpol.2021.112424.

Bistline, John E. T., Christopher Roney, Geoffrey Blanford, and David T. Young. 2023. "Impacts of Inflationary Drivers and Updated Policies on U.S. Decarbonization and Technology Transitions." White Paper 3002026229. Palo Alto, Calif.: Electric Power Research Institute. https://www.epri.com/research/products/000000003002026229.

Bistline, John E. T., Christopher W. Roney, David L. McCollum, and Geoffrey J. Blanford. 2021. "Deep Decarbonization Impacts on Electric Load Shapes and Peak Demand." *Environmental Research Letters* 16, no. 9: 094054. http://doi.org/10.1088/1748-9326/ac2197.

Bistline, John E. T., and David T. Young. 2022. "The Role of Natural Gas in Reaching Net-Zero Emissions in the Electric Sector." *Nature Communications* 13:4743. http://doi.org/10.1038/s41467-022-32468-w.

Blanford, Geoffrey J., James H. Merrick, John E. T. Bistline, and David T. Young. 2018. "Simulating Annual Variation in Load, Wind, and Solar by Representative Hour Selection." *Energy Journal* 39, no. 3: 189–212. https://doi.org/10.5547/01956574.39.3.gbla.

Böhringer, Christoph, Carolyn Fischer, Knut Einar Rosendahl, and Thomas Fox Rutherford. 2022. "Potential Impacts and Challenges of Border Carbon Adjustments." *Nature Climate Change* 12:22–29. https://doi.org/10.1038/s41558-021-01250-z.

Bolinger, Mark, Joachim Seel, Cody Warner, and Dana Robson. 2022. *Utility-Scale Solar, 2022 Edition: Empirical Trends in Deployment, Technology, Cost, Performance, PPA Pricing, and Value in the United States* [slides]. Berkeley, Calif.: Lawrence Berkeley National Laboratory. https://emp.lbl.gov/utility-scale-solar.

Borenstein, Severin, and James B. Bushnell. 2022. "Do Two Electricity Pricing Wrongs Make a Right? Cost Recovery, Externalities, and Efficiency." *American Economic Journal: Economic Policy* 14, no. 4: 80–110. https://doi.org/10.1257/pol.20190758.

Borenstein, Severin, and Ryan Kellogg. 2022. "Carbon Pricing, Clean Electricity Standards, and Clean Electricity Subsidies on the Path to Zero Emissions." Working Paper 30263. Cambridge, Mass.: National Bureau of Economic Research. https://doi.org/10.3386/w30263.

Burlig, Fiona, James B. Bushnell, David Rapson, and Catherine Wolfram. 2021. "Low Energy: Estimating Electric Vehicle Electricity Use." *American Economic Association Papers and Proceedings* 111:430–35. https://doi.org/10.1257/pandp.20211088.

Burnham, Andrew, David Gohlke, Luke Rush, Thomas Stephens, Yan Zhou, Mark A. Delucchi, and others. 2021. *Comprehensive Total Cost of Ownership Quantification for Vehicles with Different Size Classes and Powertrains*. Lemont, Ill.: Argonne National Laboratory. https://publications.anl.gov/anlpubs/2021/05/167399.pdf.

CBO (Congressional Budget Office). 2018. "How CBO Prepares Cost Estimates." Washington. https://www.cbo.gov/system/files/115th-congress-2017-2018/reports/53519-costestimates.pdf.

CBO (Congressional Budget Office). 2022. "Estimated Budgetary Effects of Public Law 117–169, to Provide for Reconciliation Pursuant to Title II of S. Con. Res. 14." Washington. https://www.cbo.gov/publication/58455.

Clausing, Kimberly A., and Catherine Wolfram. 2023. "Carbon Border Adjustments, Climate Clubs, and Subsidy Races When Climate Policies Vary." Working Paper 31310. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w31310/w31310.pdf.

Cole, Cassandra, Michael Droste, Christopher R. Knittel, Shanjun Li, and James H. Stock. 2023. "Policies for Electrifying the Light-Duty Vehicle Fleet in the United States." *American Economic Association Papers and Proceedings* 113:316–22. https://www.aeaweb.org/articles?id=10.1257/pandp.20231063.

Conejo, Antonio J., and Ramteen Sioshansi. 2018. "Rethinking Restructured Electricity Market Design: Lessons Learned and Future Needs." *International Journal of Electrical Power and Energy Systems* 98:520–30. https://doi.org/10.1016/j.ijepes.2017.12.014.

Congressional Research Service. 2022a. *Tax Provisions in the Inflation Reduction Act of 2022 (H.R. 5376)*. Washington. https://crsreports.congress.gov/product/pdf/R/R47202.

Congressional Research Service. 2022b. *Inflation Reduction Act Methane Emissions Charge: In Brief*. Washington. https://crsreports.congress.gov/product/pdf/R/R47206.

Davis, Jeffrey G., Daniel T. Kiely, George K. Haines, Isaac L. Maron, and Andre M. Smith II. 2020. "Solar and Wind Tax Credits Extended, Again." *Tax Equity Times*, December 22. https://www.taxequitytimes.com/2020/12/solar-and-wind-tax-credits-extended-again/.

DeAngelo, Julianne, Inês Azevedo, John Bistline, Leon Clarke, Gunnar Luderer, Edward Byers, and Steven J. Davis. 2021. "Energy Systems in Scenarios at Net-Zero CO_2 Emissions." *Nature Communications* 12:6096. http://doi.org/10.1038/s41467-021-26356-y.

Della Vigna, Michele, Yulia Bocharnikova, Brian Lee, Neil Mehta, Brian Singer, Enrico Chinello, and others. 2023. "Carbonomics: The Third American Energy Revolution." Goldman Sachs.

Del Negro, Marco, Julian di Giovanni, and Keshav Dogra. 2023. "Is the Green Transition Inflationary?" Staff Report 1053. Federal Reserve Bank of New York. https://www.newyorkfed.org/research/staff_reports/sr1053.

Dunkle Werner, Karl, and Stephen Jarvis. 2022. "Rate of Return Regulation Revisited." Working Paper 329. Berkeley, Calif.: Energy Institute at Haas. https://haas.berkeley.edu/wp-content/uploads/WP329.pdf.

Ela, Erik, Andrew Mills, Eric Gimon, Mike Hogan, Nicole Bouchez, Anthony Giacomoni, Hok Ng, Jim Gonzalez, and Mike DeSocio. 2021. "Electricity Market of the Future: Potential North American Designs without Fuel Costs." *IEEE Power and Energy Magazine* 19, no. 1: 41–52. https://doi.org/10.1109/MPE.2020.3033396.

Energy Innovation. 2021. "Methodology and Data Sources for the Energy Policy Simulator." San Francisco: Energy Innovation. https://energyinnovation.org/

wp-content/uploads/2021/02/Methodology-and-Data-Sources-for-the-Energy-Policy-Simulator.pdf.

EPA. 2018. *Documentation for EPA's Power Sector Modeling Platform v6 Using the Integrated Planning Model*. Washington. https://www.epa.gov/sites/default/files/2018-06/documents/epa_platform_v6_documentation_-_all_chapters_june_7_2018.pdf.

EPA. 2022. *Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances*. Washington. https://www.epa.gov/system/files/documents/2022-11/epa_scghg_report_draft_0.pdf.

EPRI. 2022. *2022 Energy System Technology Cost and Performance Summary: Market Trends and Technology Insights*. Palo Alto, Calif.: Electric Power Research Institute. https://www.epri.com/research/programs/069228/results/3002024231.

Fell, Harrison, and Joshua Linn. 2013. "Renewable Electricity Policies, Heterogeneity, and Cost Effectiveness." *Journal of Environmental Economics and Management* 66, no. 3: 688–707. https://doi.org/10.1016/j.jeem.2013.03.004.

Finkelstein, Amy, and Nathaniel Hendren. 2020. "Welfare Analysis Meets Causal Inference." *Journal of Economic Perspectives* 34, no. 4: 146–67. https://doi.org/10.1257/jep.34.4.146.

Gillingham, Kenneth, and James H. Stock. 2018. "The Cost of Reducing Greenhouse Gas Emissions." *Journal of Economic Perspectives* 32, no. 4: 53–72. https://doi.org/10.1257/jep.32.4.53.

Goulder, Lawrence H. 1995. "Environmental Taxation and the Double Dividend: A Reader's Guide." *International Tax and Public Finance* 2:157–83. https://doi.org/10.1007/bf00877495.

Goulder, Lawrence H., Marc A. C. Hafstead, GyuRim Kim, and Xianling Long. 2019. "Impacts of a Carbon Tax across US Household Income Groups: What Are the Equity-Efficiency Trade-offs?" *Journal of Public Economics* 175:44–64. https://doi.org/10.1016/j.jpubeco.2019.04.002.

Greenstone, Michael, Bogdan Mukhametkaliev, Jared Stolove, John Larsen, Ben King, Hannah Kolus, and Whitney Herndon. 2022. "Assessing the Costs and Benefits of Clean Electricity Tax Credits." Policy Memo. Chicago: Energy Policy Institute at the University of Chicago and Rhodium Group. https://epic.uchicago.edu/wp-content/uploads/2022/02/EPIC_RHO_BBBMemo.2.9.22.pdf.

Grubb, Michael, Nino David Jordan, Edgar Hertwich, Karsten Neuhoff, Kasturi Das, Kaushik Ranjan Bandyopadhyay, and others. 2022. "Carbon Leakage, Consumption, and Trade." *Annual Review of Environment and Resources* 47:753–95. https://doi.org/10.1146/annurev-environ-120820-053625.

Hirth, Lion. 2013. "The Market Value of Variable Renewables: The Effect of Solar Wind Power Variability on Their Relative Price." *Energy Economics* 38:218–36. https://doi.org/10.1016/j.eneco.2013.02.004.

Hogan, William W. 2019. "Market Design Practices: Which Ones Are Best? [In My View]" *IEEE Power and Energy Magazine* 17, no. 1: 100–104. https://doi.org/10.1109/MPE.2018.2871736.

Holland, Stephen P., Jonathan E. Hughes, and Christopher R. Knittel. 2009. "Greenhouse Gas Reductions under Low Carbon Fuel Standards?" *American Economic Journal: Economic Policy* 1, no. 1: 106–46. https://doi.org/10.1257/pol.1.1.106.

Hostert, David, Matthias Kimmel, Ian Berryman, Amar Vasdev, Nathaniel Bullard, Hugh Bromley, and others. 2022. *New Energy Outlook 2022*. BloombergNEF. https://about.bnef.com/new-energy-outlook/.

International Energy Agency. 2022. *Global EV Outlook 2022: Securing Supplies for an Electric Future*. Paris: International Energy Agency. https://iea.blob.core.windows.net/assets/ad8fb04c-4f75-42fc-973a-6e54c8a4449a/GlobalElectricVehicleOutlook2022.pdf.

International Energy Agency. 2023. *Energy Technology Perspectives 2023*. Paris: International Energy Agency. https://www.iea.org/reports/energy-technology-perspectives-2023.

Jenkins, Jesse D. 2014. "Political Economy Constraints on Carbon Pricing Policies: What Are the Implications for Economic Efficiency, Environmental Efficacy, and Climate Policy Design?" *Energy Policy* 69:467–77. https://doi.org/10.1016/j.enpol.2014.02.003.

Jenkins, Jesse D., Erin N. Mayfield, Jamil Farbes, Ryan Jones, Neha Patankar, Qingyu Xu, and Greg Schivley. 2022. *Preliminary Report: The Climate and Energy Impacts of the Inflation Reduction Act of 2022*. Princeton, N.J.: REPEAT Project. https://repeatproject.org/docs/REPEAT_IRA_Prelminary_Report_2022-08-04.pdf.

Jiang, Betty, Maheep Mandloi, Richard Carlson, Matt Hope, Mike Ziffer, Nick Campanella, and others. 2022. *US Inflation Reduction Act: A Tipping Point in Climate Action*. Zurich: Credit Suisse. https://www.credit-suisse.com/about-us-news/en/articles/news-and-expertise/us-inflation-reduction-act-a-catalyst-for-climate-action-202211.html.

Joint Committee on Taxation (JCT). 2019. "The Joint Committee on Taxation Revenue Estimating Process." Washington. https://www.jct.gov/CMSPages/GetFile.aspx?guid=5a50d049-0af4-4658-a174-d35777b4418e.

Joskow, Paul L. 2011. "Comparing the Costs of Intermittent and Dispatchable Electricity Generating Technologies." *American Economic Review* 101, no. 3: 238–41. https://doi.org/10.1257/aer.101.3.238.

Juhász, Réka. 2018. "Temporary Protection and Technology Adoption: Evidence from the Napoleonic Blockade." *American Economic Review* 108, no. 11: 3339–76. https://doi.org/10.1257/aer.20151730.

Kotchen, Matthew J., and Erin T. Mansur. 2014. "How Stringent Are the US EPA's Proposed Carbon Pollution Standards for New Power Plants?" *Review of Environmental Economics and Policy* 8, no. 2: 290–306. https://doi.org/10.1093/reep/reu003.

Krosnick, Jon A., and Bo MacInnis. 2020. *Climate Insights 2020: Overall Trends*. Washington: Resources for the Future. https://www.rff.org/publications/reports/climateinsights2020/.

Larsen, John, Ben King, Hannah Kolus, Naveen Dasari, Galen Hiltbrand, and Whitney Herndon. 2022. *A Turning Point for US Climate Progress: Assessing the Climate and Clean Energy Provisions in the Inflation Reduction Act*. New York: Rhodium Group. https://rhg.com/research/climate-clean-energy-inflation-reduction-act/.

Levin, Amanda, and Jacqueline Ennis. 2022. "Clean Electricity Tax Credits in the Inflation Reduction Act Will Reduce Emissions, Grow Jobs, and Lower Bills." Issue Brief. New York: Natural Resources Defense Council. https://www.nrdc.org/resources/clean-electricity-tax-credits-inflation-reduction-act-will-reduce-emissions-grow-jobs-and.

Linn, Joshua. 2022. "Balancing Equity and Effectiveness for Electric Vehicle Subsidies." Working Paper 22-7. Washington: Resources for the Future. https://www.rff.org/publications/working-papers/balancing-equity-and-effectiveness-for-electric-vehicle-subsidies/.

Lonergan, Katherine E., Florian Egli, Sebastian Osorio, Giovanni Sansavini, Michael Pahle, Tobias S. Schmidt, and Bjarne Steffen. 2023. "Improving the Representation of Cost of Capital in Energy System Models." *Joule* 7, no. 3: 469–83. https://doi.org/10.1016/j.joule.2023.02.004.

Mahajan, Megan, Olivia Ashmoore, Jeffrey Rissman, Robbie Orvis, and Anand Gopal. 2022. *Modeling the Inflation Reduction Act Using the Energy Policy Simulator*. San Francisco: Energy Innovation. https://energyinnovation.org/wp-content/uploads/2022/08/Modeling-the-Inflation-Reduction-Act-with-the-US-Energy-Policy-Simulator_August.pdf.

Malik, Naureen S. 2022. "Negative Power Prices? Blame the US Grid for Stranding Renewable Energy." Bloomberg, August 30. https://www.bloomberg.com/news/articles/2022-08-30/trapped-renewable-energy-sends-us-power-prices-below-zero#xj4y7vzkg.

Mays, Jacob, David P. Morton, and Richard P. O'Neill. 2019. "Asymmetric Risk and Fuel Neutrality in Electricity Capacity Markets." *Nature Energy* 4:948–56. https://doi.org/10.1038/s41560-019-0476-1.

McArdle, Paul, and Elesia Fasching. 2022. "Carbon Intensity of U.S. Power Generation Continues to Fall but Varies Widely by State." US Energy Information Administration, September 13. https://www.eia.gov/todayinenergy/detail.php?id=53819.

McKerracher, Colin, Aleksandra O'Donovan, Nikolas Soulopoulos, Andrew Grant, Jinghong Lyu, Siyi Mi, and others. 2022. *Electric Vehicle Outlook 2023*. BloombergNEF. https://about.bnef.com/electric-vehicle-outlook/.

O'Boyle, Mike, Dan Esposito, and Michelle Solomon. 2022. *Implementing the Inflation Reduction Act: A Roadmap for State Electricity Policy*. San Francisco: Energy Innovation. https://energyinnovation.org/wp-content/uploads/2022/10/Implementing-the-Inflation-Reduction-Act-A-Roadmap-For-State-Policy-1.pdf.

OECD. 2022. *Pricing Greenhouse Gas Emissions: Turning Climate Targets into Climate Action*. Paris: Organisation for Economic Cooperation and Development. https://doi.org/10.1787/e9778969-en.

Paul, Anthony, Karen Palmer, and Matthew Woerman. 2015. "Incentives, Margins, and Cost Effectiveness in Comprehensive Climate Policy for the Power Sector." *Climate Change Economics* 6, no. 4: 1550016. https://doi.org/10.1142/S2010007815500165.

Peñasco, Cristina, Laura Díaz Anadón, and Elena Verdolini. 2021. "Systematic Review of the Outcomes and Trade-offs of Ten Types of Decarbonization Policy Instruments." *Nature Climate Change* 11:257–65. http://doi.org/10.1038/s41558-020-00971-x.

Popp, David, Richard G. Newell, and Adam B. Jaffe. 2010. "Energy, the Environment, and Technological Change." In *Handbook of the Economics of Innovation, Volume 2*, edited by Bronwyn H. Hall and Nathan Rosenberg. Amsterdam: North-Holland. https://doi.org/10.1016/S0169-7218(10)02005-8.

Raimi, Daniel, and Sophie Pesek. 2022. *What Is an "Energy Community"? Alternative Approaches for Geographically Targeted Energy Policy*. Washington: Resources for the Future. https://media.rff.org/documents/Report_22-12_AxXwJqy.pdf.

Rennert, Kevin, Frank Errickson, Brian C. Prest, Lisa Rennels, Richard G. Newell, William Pizer, and others. 2022. "Comprehensive Evidence Implies a Higher Social Cost of CO_2." *Nature* 610:687–92. http://doi.org/10.1038/s41586-022-05224-9.

Rivera, Alfredo, Ben King, John Larsen, and Kate Larsen. 2023. "Preliminary US Greenhouse Gas Emissions Estimates for 2022." New York: Rhodium Group. https://rhg.com/research/us-greenhouse-gas-emissions-2022/.

Roy, Nicholas, Dallas Burtraw, and Kevin Rennert. 2021. "Cost Analysis and Emissions Projections under Power Sector Proposals in Reconciliation." Issue Brief 21-15. Washington: Resources for the Future. https://media.rff.org/documents/IB_21-15_cTjoJFj.pdf.

Roy, Nicholas, Dallas Burtraw, and Kevin Rennert. 2022. "Retail Electricity Rates under the Inflation Reduction Act of 2022." Issue Brief 22-07. Washington: Resources for the Future. https://media.rff.org/documents/IB_22-07_HcKDycO.pdf.

Roy, Nicholas, Maya Domeshek, Dallas Burtraw, Karen Palmer, Kevin Rennert, Jhih-Shyang Shih, and Seth Villanueva. 2022. *Beyond Clean Energy: The Financial Incidence and Health Effects of the IRA*. Washington: Resources for the Future. https://media.rff.org/documents/Report_22-11_v5.pdf.

Seel, Joachim, Dev Millstein, Andrew Mills, Mark Bolinger, and Ryan Wiser. 2021. "Plentiful Electricity Turns Wholesale Prices Negative." *Advances in Applied Energy* 4:100073. https://doi.org/10.1016/j.adapen.2021.100073.

Senate Democrats. 2022. "Joint Statement from Leader Schumer and Senator Manchin Announcing Agreement to Add the Inflation Reduction Act of 2022 to the FY2022 Budget Reconciliation Bill and Vote in Senate Next Week." Washington. https://www.democrats.senate.gov/newsroom/press-releases/senate-majority-leader-chuck-schumer-d-ny-and-sen-joe-manchin-d-wv-on-wednesday-announced-that-they-have-struck-a-long-awaited-deal-on-legislation-that-aims-to-reform-the-tax-code-fight-climate-change-and-cut-health-care-costs.

Shindell, Drew, Muye Ru, Yuqiang Zhang, Karl Seltzer, Greg Faluvegi, Larissa Nazarenko, and others. 2021. "Temporal and Spatial Distribution of Health, Labor, and Crop Benefits of Climate Change Mitigation in the United States." *Proceedings of the National Academy of Sciences* 118, no. 46: e2104061118. http://doi.org/10.1073/pnas.2104061118.

Slowik, Peter, Stephanie Searle, Hussein Basma, Josh Miller, Yuanrong Zhou, Felipe Rodríguez, and others. 2023. "Analyzing the Impact of the Inflation Reduction Act on Electric Vehicle Uptake in the United States." White Paper. San Francisco: Energy Innovation and the International Council on Clean Transportation. https://theicct.org/wp-content/uploads/2023/01/ira-impact-evs-us-jan23-2.pdf.

Stock, James H., and Daniel N. Stuart. 2021. "Robust Decarbonization of the US Power Sector: Policy Options." Working Paper 28677. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28677.

Timilsina, Govinda R. 2022. "Carbon Taxes." *Journal of Economic Literature* 60, no. 4: 1456–502. http://doi.org/10.1257/jel.20211560.

US Department of the Treasury. 2023a. "Treasury Releases Proposed Guidance on New Clean Vehicle Credit to Lower Costs for Consumers, Build U.S. Industrial Base, Strengthen Supply Chains." Press release, March 31. https://home.treasury.gov/news/press-releases/jy1379.

US Department of the Treasury. 2023b. "Treasury Releases Guidance to Drive Investment to Coal Communities." Press release, April 4. https://home.treasury.gov/news/press-releases/jy1383.

US Energy Information Administration. 2022. *Annual Energy Outlook 2022: Alternative Policies for Carbon Fee Cases*. Washington: Department of Energy. https://www.eia.gov/outlooks/aeo/IIF_carbonfee/pdf/carbon_fee_analysis.pdf.

US Energy Information Administration. 2023. *Annual Energy Outlook 2023: Issues in Focus—Inflation Reduction Act Cases in the AEO2023*. Washington: Department of Energy. https://www.eia.gov/outlooks/aeo/IIF_IRA/pdf/IRA_IIF.pdf.

van den Bergh, Jeroen, and Ivan Savin. 2021. "Impact of Carbon Pricing on Low-Carbon Innovation and Deep Decarbonisation: Controversies and Path Forward." *Environmental and Resource Economics* 80:705–15. http://doi.org/10.1007/s10640-021-00594-6.

Wang, Seaver, Zeke Hausfather, Steven Davis, Juzel Lloyd, Erik B. Olsen, Lauren Liebermann, Guido D. Núñez-Mujica, and Jameson McBride. 2023. "Future Demand for Electricity Generation Materials under Different Climate Mitigation Scenarios." *Joule* 7, no. 2: 309–32. http://doi.org/10.1016/j.joule.2023.01.001.

Way, Rupert, Matthew C. Ives, Penny Mealy, and J. Doyne Farmer. 2022. "Empirically Grounded Technology Forecasts and the Energy Transition." *Joule* 6, no. 9: 2057–82. https://doi.org/10.1016/j.joule.2022.08.009.

White House. 2022. "Remarks by President Biden on the Passage of H.R. 5376, the Inflation Reduction Act of 2022." White House South Lawn, September 13. https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/09/13/remarks-by-president-biden-on-the-passage-of-h-r-5376-the-inflation-reduction-act-of-2022/

White House. 2023a. *Building a Clean Energy Economy: A Guidebook to the Inflation Reduction Act's Investments in Clean Energy and Climate Action*. Washington. https://www.whitehouse.gov/wp-content/uploads/2022/12/Inflation-Reduction-Act-Guidebook.pdf.

White House. 2023b. "Treasury Releases New Guidance to Drive Clean Energy Investment to America's Energy Communities." Press release, April 4. https://www.whitehouse.gov/cleanenergy/clean-energy-updates/2023/04/04/treasury-releases-new-guidance-to-drive-clean-energy-investment-to-americas-energy-communities/.

Williams, Mike, Marina Zhavoronkov, and David Madland. 2022. "The Inflation Reduction Act Provides Pathways to High-Quality Jobs." Center for American Progress. https://www.americanprogress.org/article/the-inflation-reduction-act-provides-pathways-to-high-quality-jobs/.

Wiser, Ryan, Mark Bolinger, Ben Hoen, Dev Millstein, Joe Rand, Galen Barbose, Naïm Darghouth, Will Gorman, Seongeun Jeong, and Ben Paulos. 2022. *Land-Based Wind Market Report: 2022 Edition*. Berkeley, Calif.: Lawrence Berkeley National Laboratory. https://emp.lbl.gov/sites/default/files/2022_land_based_wind_market_report.pdf.

Wright, T. P. 1936. "Factors Affecting the Cost of Airplanes." *Journal of the Aeronautical Sciences* 3, no. 4: 122–28. http://doi.org/10.2514/8.155.

Xu, Qingyu, Neha Patankar, Michael Lau, Chuan Zhang, and Jesse D. Jenkins. 2022. *Cleaner, Faster, Cheaper: Impacts of the Inflation Reduction Act and a Blueprint for Rapid Decarbonization in the PJM Interconnection*. Geneva: Zenodo. https://doi.org/10.5281/zenodo.7429042.

Young, David, and John Bistline. 2018. "The Costs and Value of Renewable Portfolio Standards in Meeting Decarbonization Goals." *Energy Economics* 73:337–51. https://doi.org/10.1016/j.eneco.2018.04.017.

Zhao, Alicia, Shannon Kennedy, Kowan O'Keefe, Maria Borreo, Kyle Clark-Sutton, Ryna Cui, and others. 2022. *An "All-In" Pathway to 2030: The Beyond 50 Scenario*. College Park, Md.: Center for Global Sustainability, University of Maryland and America Is All In. https://cgs.umd.edu/sites/default/files/2022-11/CGS_UMD_Beyond50report_Nov2022.pdf.

Comments and Discussion

COMMENT BY

JASON FURMAN I do not remember ever seeing a more sophisticated after-action report on a piece of legislation that combines public finance, macroeconomics, and statistics on the environment, together with lots of institutional details. Which makes me even more guilty that my main reaction is to list everything I wanted to know that did not fit into the paper—and Bistline, Mehrotra, and Wolfram would probably agree with most of my wish list. So my comments are mostly about using their paper as a springboard for some broader observations about the Inflation Reduction Act (IRA) and also different approaches to climate change. My comment makes four points.

1. WILL THE INFLATION REDUCTION ACT REDUCE THE DEFICIT? Bistline, Mehrotra, and Wolfram provide updated estimates of the cost of the climate provisions of the IRA that are considerably higher than the initial Congressional Budget Office (CBO) score of the legislation. These higher numbers are generally in line with a number of other ex post estimates made by the climate analytics community. But the authors do not provide any bottom line estimates of the cost of the full legislation because they do not reestimate the other pieces outside the climate domain.

To provide greater context, the first CBO score projected that the initial version of the IRA would reduce or have minor effects on the deficit every single year. The legislation was changed in the process of the negotiation, resulting in a CBO score that showed it raising the deficit slightly in the first five years but cutting it over the full ten-year window, as shown in table 1.

After the law passed, the administration scored it for the PAYGO (pay as you go) scorecard. The administration's estimate, including its estimate

Brookings Papers on Economic Activity, Spring 2023: 158–182 © 2024 The Brookings Institution.

Table 1. Scores for the Inflation Reduction Act

	2022–2026	*2027–2031*	*2022–2031*
Nominal ($ billions)			
CBO	26	–264	–238
Administration	151	–341	–190
CBO plus authors' scores	248	24	272
CBO plus authors' scores and higher raisers	202	–225	–23
Percentage of GDP (%)			
CBO	0.0	–0.2	–0.1
Administration	0.1	–0.2	–0.1
CBO plus authors' scores	0.2	0.0	0.1
CBO plus authors' scores and higher raisers	0.1	–0.1	0.0

Source: Author's calculations based on Congressional Budget Office, Office of Management and Budget, and data provided by Bistline, Mehrotra, and Wolfram.

of the unscored revenue from tax reform, was that the legislation would raise the deficit slightly less over the ten-year window—with larger deficit increases in the first five years and larger deficit reduction in the second five years. The administration did not provide any detail about the components of its score or why it differed from the CBO, but it is reasonable to guess that the difference at least partly reflects the administration's higher score for the energy provisions partly offset by a larger score for the tax offsets.

If you combine the CBO score of the non-climate provisions with the authors' score of the climate provisions, the result is that the legislation raises the deficit over ten years by $272 billion, with a larger deficit increase in the first five years than in the last five years.

The CBO, however, likely underestimated how much revenue the tax increases included in the legislation would raise. First, its score of the tax enforcement provisions explicitly did not include any effect of the legislation on deterrence because of their scorekeeping practices, not because they argued this would be zero. The administration's estimate of the enforcement savings was higher. Second, the nominal corporate income baseline has been much higher than the assumption underlying the CBO score, which was set a year and a half before the legislation passed. Just shifting corporate revenues up proportionally to the higher baseline results in additional savings. Finally, initial indications are that behavior is less sensitive to the buyback tax than expected, leading that measure to raise more revenue. Combining the authors' climate provisions, plausibly rescored tax provisions, and the CBO estimates for everything else, the result is that the legislation costs $202 billion in the first five years and raises $225 billion in the second five years—for a ten-year budgetary savings of $23 billion.

Overall, especially if you look at the numbers as a fraction of GDP, the deficit impact of the legislation basically rounds to zero—and then we can argue about the second, third, and fourth decimal places but we will never truly know.

2. WILL THE INFLATION REDUCTION ACT AFFECT THE MACROECONOMY? Much of the authors' paper is a sophisticated macroeconomic analysis of the IRA. This analysis convinces me that the macroeconomics of the IRA are not especially important. Figure 17 in the paper shows their major macroeconomic results. These magnitudes can be put into perspective by looking at both the baseline and the changes, as shown in figure 1, using numbers provided by the authors in March 2023. Here you can see that the impact of the IRA is essentially imperceptible.

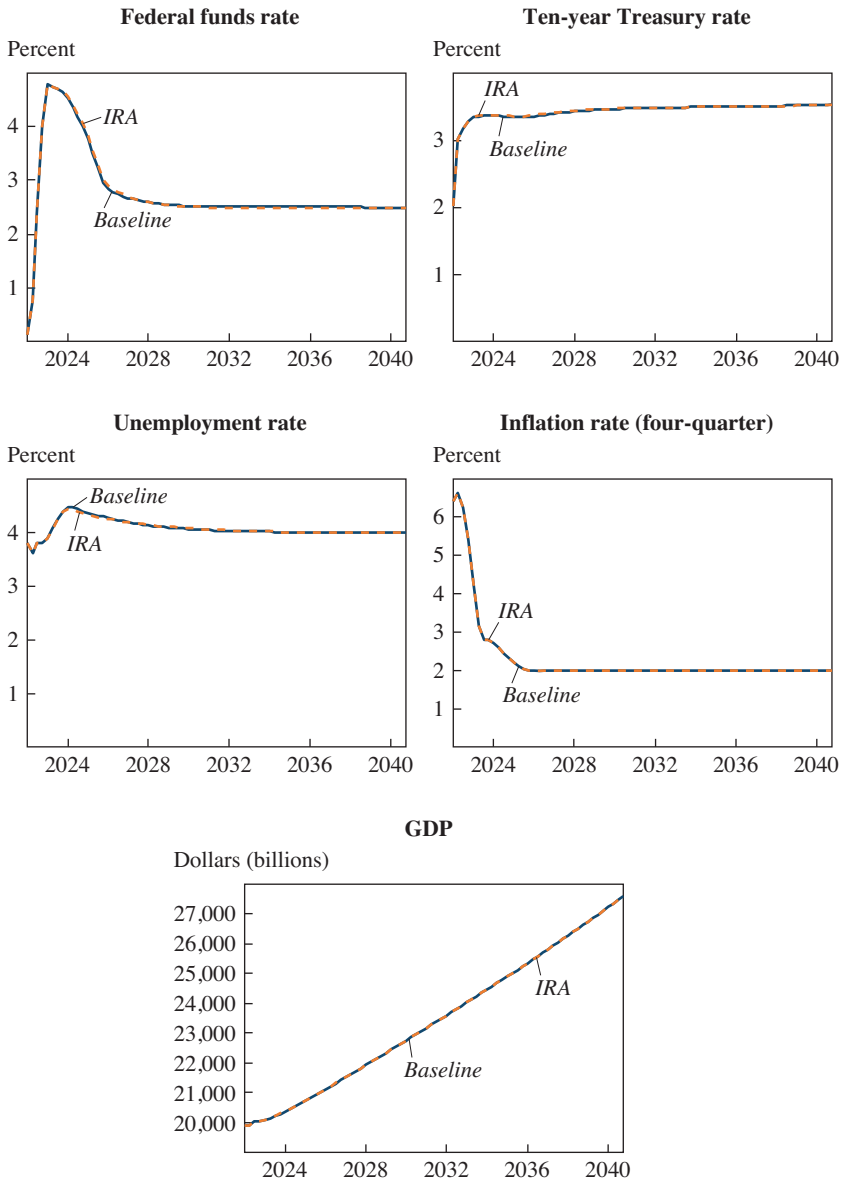
Note these estimates include only the climate parts of the legislation. To close their model, the authors assume lump sum financing for the rest. As a result, they are analyzing a situation in which the user cost of capital is reduced.

Of course, the rest of the legislation did not involve lump sum financing. In fact, the corporate tax increases included in the legislation have the opposite effect, raising the user cost of capital. Thus the legislation, as a whole, might not actually subsidize capital; instead it might subsidize capital in the climate sector but raise the bar for it outside the climate sector.

The authors' estimates also do not consider the Keynesian impact of the legislation. Using a set of multipliers consistent with what the CBO has proposed in the past, figure 2 shows the macroeconomic impact of the legislation on nominal GDP—with the breakout between prices and quantities ambiguous and dependent on the cyclical state of the economy. Moreover, this assumes no Federal Reserve reaction, so the estimates become increasingly less relevant over the ten-year window. Regardless, the most important result is that this effect, too, is relatively small.

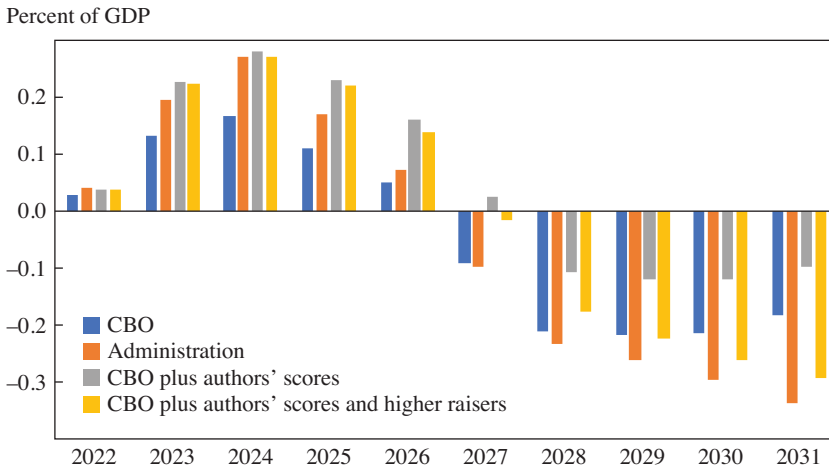
There are other macroeconomic effects that also do not show up in the authors' analysis, which could be just as big as the ones that do. The legislation will have large distributional effects (discussed in the next section) that could feed back into consumption, interest rates, and growth. The sectoral shifts that it causes could be first order, potentially raising unemployment over time by worsening job matching. Induced changes in other countries through technological spillovers and change could be important. Finally, the authors do not attempt to analyze any macroeconomic impact of the changes in climate that result from this legislation—although those are likely to be small.

Figure 1. Macroeconomic Effects of the Inflation Reduction Act



Source: Data provided by Bistline, Mehrotra, and Wolfram.

Figure 2. Effect of Inflation Reduction Act on Nominal GDP



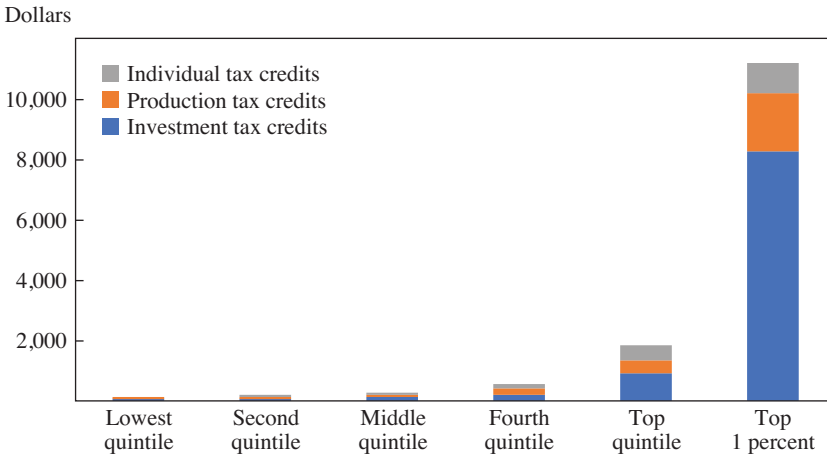
Source: Author's calculations based on Congressional Budget Office, Office of Management and Budget, and data provided by Bistline, Mehrotra, and Wolfram.

3. WILL THE INFLATION REDUCTION ACT AFFECT THE DISTRIBUTION OF AFTER-TAX INCOME? The distributional impact of the IRA is considerably more important than the macroeconomic impact. The authors' Federal Reserve Board US model (FRB/US) simulation finds that the legislation raises GDP by $79 billion through 2031. In contrast, the US Regional Economy, Greenhouse Gas, and Energy (US-REGEN) model's central scenario has $781 billion in total tax credits through 2031. How the latter are distributed could be considerably more important than the GDP effect.

In fact, these tax subsidies are roughly the same size as the earned income tax credit (EITC), the Supplemental Nutrition Assistance Program (SNAP), Supplemental Security Income (SSI), or the health insurance premium tax credit. Nevertheless, the distribution of the tax credits has received almost no discussion—in contrast to the endless discussions about the distribution of a carbon tax.

So what is the incidence of the subsidies? The authors say, I think plausibly, that most of them are passed on to consumers. But about 60 percent of those consumers are commercial and industrial users of electricity so it also matters what those consumers do with the savings. In some ways the subsidies are more like a corporate tax cut. In addition, the incidence also depends on regulatory rules as well as the comparative elasticity of supply

Figure 3. Reduction in Average Taxes from Climate Tax Provisions of the Inflation Reduction Act, 2027



Source: Author's calculations based on data from the Tax Policy Center.

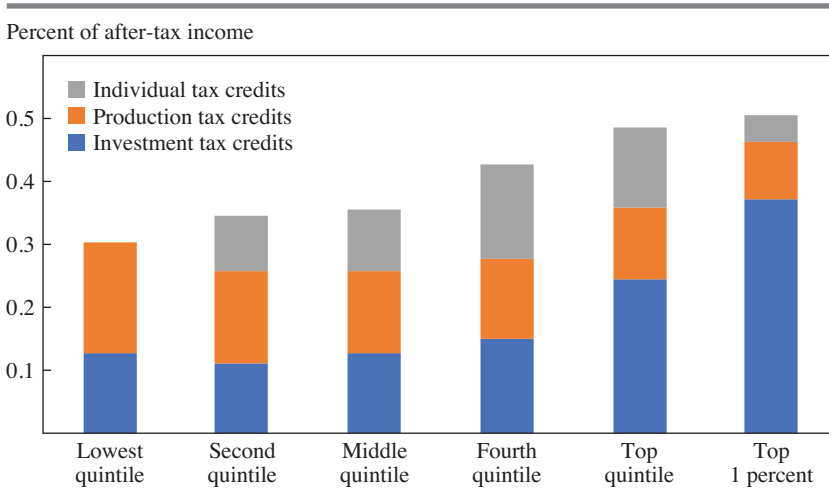
and demand. And at least for some users, like industrial, the demand is not perfectly inelastic. Moreover, demand is more elastic over longer periods of time—suggesting that more of the incidence of the tax cuts may fall on producers over time.

To get a rough indication of the distributional impact of the legislation I took some unpublished tables that the Tax Policy Center provided to me and readjusted them to match the authors' estimates of the total cost in dollars. The breakdown of the resulting tax cuts are shown in figure 3, with the large majority going to the top 1 percent.

The best way to assess the progressivity is to show those tax changes as a share of after-tax income, as is done in figure 4. This shows that overall the climate provisions in the legislation may have been slightly regressive—providing larger gains to higher-income households than to middle-income households. Note also that the average increase in after-tax income from the incidence of the tax changes is about 0.4 percent, which is about twenty times larger than the approximately 0.02 percent increase in GDP in 2027.

As an aside, a carbon tax raises taxes on people. As a result they have gotten much more scrutiny than these sorts of industrial policies that give money to companies. If you are doing industrial policy, you are compensating

Figure 4. Change in After-Tax Income from Climate Tax Provisions of the Inflation Reduction Act, 2027



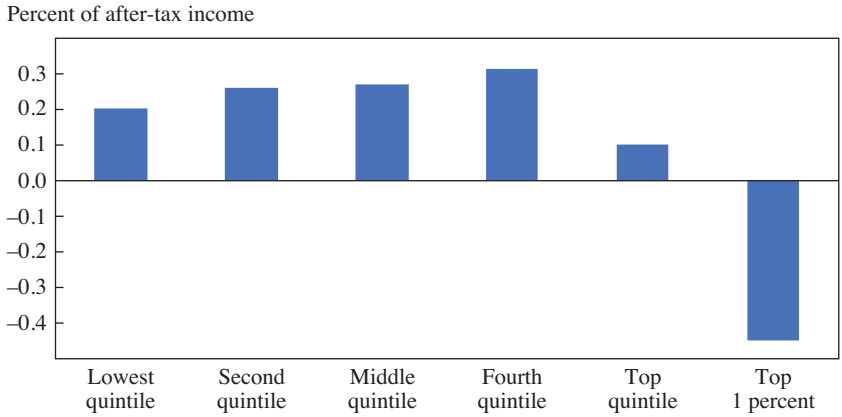
Source: Author's calculations based on data from the Tax Policy Center.

people through businesses, you are giving money to businesses and hoping that money eventually gets to people. The idea that you can count on that in any sort of way is much less certain and deserves more attention. These exact numbers for the distribution may be wrong; my point is more that these issues require much more attention—even relative to the macroeconomic impact.

Note that this point is about the climate provisions only. The legislation as a whole was almost certainly progressive when the highly progressive tax raisers are taken into account. This is shown in figure 5.

4. CARBON TAXES VERSUS CLEAN ENERGY SUBSIDIES The authors do not break out the consequences of the different subsidies in the IRA or which have high bang for the buck, which have low, and what the consequences of the unevenness are. Reading between the lines in the paper, it appears as if the authors think the investment tax credit and production tax credit are very valuable, possibly almost as good as a carbon tax. But the electric vehicle subsidies have a substantial inframarginal cost. Some provisions have close to zero or even a negative impact on carbon emissions, like domestic content bonuses and manufacturing subsidies. In addition, other aspects of the legislation, like bonuses for paying prevailing wages, have an ambiguous and possibly negative impact on carbon emissions.

Figure 5. Change in After-Tax Income from the Inflation Reduction Act, 2027



Source: Author's calculations based on data from the Tax Policy Center.

Table 2. Change in Generation Share from 2021 to 2035

	Percentage points	
	IRA	*Carbon tax*
Coal	–14	–18
Natural gas	–21	–5
Coal CCS	3	0
Wind and solar	28	19
Other	4	4
CO_2 (% drop from 2005)	68%	68%
Abatement cost ($/$t$ – CO_2)	$45–61	$10

Source: Author's calculations based on table 3 in the paper.

Taking the legislation as a whole, it is possible to compare it to a carbon tax. According to the authors, the IRA gets the same amount of carbon reductions as a $10 carbon tax, both resulting in a 68 percent drop in carbon usage in 2035 relative to 2005. The way in which they get these reductions, however, is very different, as shown in table 2.

To the degree that the carbon tax is the gold standard—with those changes reflecting an optimal Pigouvian tax that internalizes the externalities—then the differences in the IRA reflect ways in which it results in an unnecessarily high social cost of emissions reductions. In particular, a carbon tax gets the same emissions reductions more cheaply by relying more on shifting from coal to natural gas. By comparison, the IRA actually results in more coal

use (because coal is not especially penalized as compared to natural gas) but makes up for this with even larger wind and solar. It is possible, however, that the IRA approach will be better than a carbon tax in leading to more induced innovation.

Clean energy subsidies and carbon taxes can be equivalent if demand is inelastic on every timescale and innovation is unresponsive to price. In reality these are both unlikely to be true, and in particular, the fact that the US-REGEN model predicts the legislation will lower electricity prices by about 13 percent in 2050 makes me concerned that it will reduce innovations for greater energy efficiency.

As discussed above, clean energy subsidies raise issues regarding the distribution of their benefits. To the degree that they are mildly regressive this creates a headwind for the progressivity of the overall legislation. In contrast, carbon taxes raise revenues that can be explicitly allocated to achieve almost any distributional goal.

Finally, carbon taxes operate on multiple margins. For example, in the vehicle sector they do not just affect the choice of vehicle but also the amount of driving. In addition, they operate across sectors—not just in the energy sector.

Of course policymakers considering the IRA could not compare it to the almighty (a carbon tax) but had to compare it to the alternative (do nothing). And the authors make a convincing case that it is considerably better than that with a $45–61 per ton abatement cost that is considerably below a plausible social cost of carbon.

SUMMARY AND BIG QUESTIONS GOING FORWARD The net economic impact of the IRA, especially when the offsets (which are beyond the scope of this paper) are included, was approximately zero and very uncertain. The macroeconomic impact was very small. The distributional impact was an order of magnitude larger and possibly mildly regressive. Relative to a carbon price, the actual subsidy policy may be far from optimal, leading to sustained electricity price declines and distributional impacts—but it is still better than nothing.

Going forward, the most important questions concern what the next steps are. The IRA is equivalent to a $10 per ton carbon tax, so much more will need to be done to meet the US net zero commitments or the Paris commitment, or even just to have the optimal Pigouvian tax. The IRA, however, may not be easily scalable—and certainly could not be scaled ten times, which is what would be needed to approximate the social cost of carbon. Are there other approaches, other than carbon pricing, that could work at scale? And if so, what are they—and what are the most efficient ways to scale them up?

COMMENT BY

KENNETH GILLINGHAM The global mean surface temperature has increased by nearly 2 degrees Fahrenheit since 1880, and the rate of warming has doubled since 1981 to 0.32 degrees Fahrenheit per decade (Lindsey and Dahlman 2023). Clearly the climate is changing, and the science is settled that greenhouse gas emissions from human activities are far and away the dominant cause (IPCC 2023). The Inflation Reduction Act (IRA) of 2022 made the largest outlay toward greenhouse gas emission reductions in the history of the United States as part of a larger package that included lowering prescription drug prices. One of the major goals of the IRA was the reduce the US government fiscal deficit, for the expectation was that the savings from the non-climate spending measures would easily offset the spending from climate measures.

Bistline, Mehrotra, and Wolfram make a convincing argument that the uptake of the climate incentives may be several times greater than expected, implying that the fiscal costs of the climate provisions in the IRA may be much larger than projected. In fact, the numbers in the paper suggest that the entire bill might possibly be in the red when it comes to the fiscal impact. This finding aligns with a few other recent studies and suggests that the official scoring of the IRA could be quite far off (Jenkins and others 2022; O'Boyle, Esposito, and Solomon 2022; Roy and others 2022). From a climate perspective, this could be seen as a reason to rejoice, as it implies greater emission reductions. From a fiscal perspective, this could be seen as quite problematic. It also raises questions about the political durability of the IRA if the costs substantially exceed expectations.

Bistline, Mehrotra, and Wolfram's contributions go far beyond making the point that the fiscal cost of the IRA may exceed expectations. In this comment, I delve into what I consider the major contributions of the paper to be. I then discuss five key points that I believe emerge from the work that indicate how the IRA will be shaping emissions and the fiscal budget in the next decade.

CONTRIBUTIONS Bistline, Mehrotra, and Wolfram perform a set of runs with a large-scale energy model of the United States, US Regional Economy, Greenhouse Gas, and Energy (US-REGEN). This is an extensively developed model created and run by the Electric Power Research Institute. My view of the model is that it is exceptionally careful in the modeling of the electricity system, right in the middle of the pack in its assumptions for electric vehicle market share, and perhaps somewhat optimistic in the modeling of electrifying building end uses. No model is perfect, but US-REGEN provides a useful set of results that can provide real insights

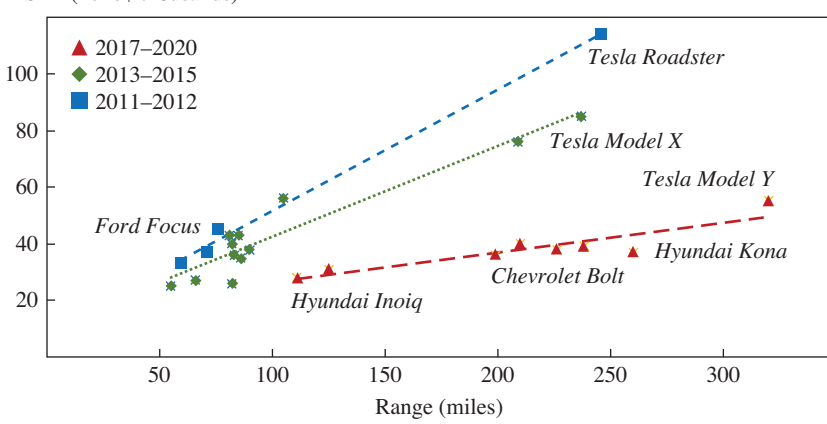
into the nature of the energy system. The results of the US-REGEN model suggest that the IRA can lead to economy-wide reductions of carbon dioxide emissions from 2005 levels of 35 percent by 2030, an estimated emission reduction in line with other studies and only modestly lower than the announced reductions of 40 percent by 2030 (Senate Democrats 2022).

Total fiscal cost and electric vehicles. The central case US-REGEN results suggest is that the total undiscounted fiscal cost of the IRA by 2031 could be approximately $780 billion. This is roughly double the $392 billion scoring from the Congressional Budget Office (CBO 2022), which used inputs from the Joint Committee on Taxation (JCT). Of course, the IRA is a bundle of different tax credits. The biggest difference between the two estimates stems from the treatment of electric vehicle tax credits. The CBO/JCT estimate for the cumulative cost of the electric vehicle tax credit through 2031 is $14 billion, which followed based on an assumption that the take-up of these tax credits would be low due to certain provisions. Specifically, the electric vehicle tax credits in the IRA can be as high as $7,500, but to obtain the full tax credit the batteries and critical minerals are required to be at least partly domestically sourced. Moreover, there are income thresholds that make high-income households (often the households that buy new cars) ineligible for taking the tax credits and price caps so that the most expensive electric vehicles are not eligible. There are also used vehicle tax credits with income thresholds and a price cap of $25,000 for the used vehicles. US-REGEN makes very different assumptions regarding these provisions.

Rather than the $14 billion CBO/JCT score, the US-REGEN score for the total fiscal costs of clean vehicle credits is $390 billion. This is a massive difference! Indeed, it accounts for a large fraction of the differences between the two scores. The US-REGEN central case results suggest an electric vehicle market share in 2030 of 44 percent (increased by the IRA from 32 percent). Recent work by Cole and others (2023), as well as a report by Goldman Sachs (2023), also suggests substantially higher cumulative spending on electric vehicles. Cole and others (2023) estimate cumulative federal tax expenditures for electric vehicles of approximately $450 billion by 2030.

Are these much larger cumulative expenditures on electric vehicles possible? They definitely are. Recent work in the literature, including Gillingham and others (2023) and Forsythe and others (2023), suggests that there is substantial unmet demand for electric vehicles. The National Academies of Sciences, Engineering, and Medicine (2021) highlight the remarkable innovations that have occurred over the past decade in electric

Figure 1. Electric Vehicles: Price versus Range



Source: Gillingham and Stock (2018), updated to reflect newer data.

Note: Scatterplot of manufacturer suggested retail price (MSRP) when introduced versus vehicle range (in miles), grouped by vehicle release year. Trend lines show linear regression fit.

vehicle technology, while also projecting continued improvements in the technology going forward. This substantial innovation can also be seen in figure 1, which plots vehicle price against vehicle range, showing a steady march downward in the cost of vehicles along with an increase in range over time.

With the continued innovation projected by the National Academies, a 44 percent market share of electric vehicles in the new vehicle fleet by 2031 appears completely reasonable. It is substantially less than would be needed if the US Environmental Protection Agency (EPA) light-duty vehicle greenhouse gas standards are finalized at the currently proposed levels—which EPA modeling suggests could be met at 67 percent electric vehicles by 2032.[1]

However, just because it is very possible that the electric vehicle market share will be at 44 percent by 2031 does not necessarily mean that all of the new vehicles will receive the full tax incentives. The US Treasury has thus far read the language of the IRA in a way that permits more vehicles to be eligible. For example, there is a "commercial vehicle" provision that

1. Environmental Protection Agency, "Proposed Rule: Multi-Pollutant Emissions Standards for Model Years 2027 and Later Light-Duty and Medium-Duty Vehicles," https://www.epa.gov/regulations-emissions-vehicles-and-engines/proposed-rule-multi-pollutant-emissions-standards-model.

permits fleets to purchase electric vehicles without the domestic content provision. This clause is being used by some automakers (most notably Kia and Hyundai) to offer much lower lease rates by having the leasing company take the tax credit (Schaffner 2023). It is not clear how the domestic content provisions will be interpreted going forward, and it will depend in part on future administrations. Moreover, there is substantial uncertainty in the evolution of the vehicle market, let alone the evolution of electric vehicle technology. The CBO/JCT score takes a notably more pessimistic view of growth in electric vehicle market share and eligibility of vehicles and vehicle buyers for tax credits.

There are other differences between the assumptions of US-REGEN and the CBO/JCT scoring, such as the likelihood that the tax credits for carbon capture and sequestration (45Q credits) will be used, but these are swamped by the difference in the electric vehicle tax credits. There is substantial uncertainty in whether carbon capture and sequestration will ever be economically viable, with the cost currently remaining quite high. But the new EPA-proposed rule for power plant greenhouse gas emissions would put immense pressure on fossil fuel–powered electricity generating units to add carbon capture and sequestration technology should it be finalized. Thus, there is reason to believe that the US-REGEN estimates of the growth of carbon capture and sequestration could really come to pass.

Cost-effectiveness of the IRA. Bistline, Mehrotra, and Wolfram make the ambitious effort to estimate the cost-effectiveness of the IRA power sector climate provisions. They take an administrative cost perspective, which divides the total IRA expenditure by the total emission reductions from the legislation. Note that this perspective is not the same as a social welfare perspective. Indeed, much of the expenditure can be thought of as transfers from taxpayers to recipients of the tax credits. For a social welfare perspective, the social cost of public funds would have to be applied to the expenditures to calculate the deadweight loss from raising the tax revenue. If the tax revenue was raised by correcting externalities, there could even be no social cost of public funds. Yet, the administrative cost perspective is the standard perspective taken in the literature, and it is useful to consider this perspective for comparison purposes.

The primary cost-effectiveness result in the paper may be surprising to some. The authors calculate several different estimates and restrict the analysis to looking only at impacts in the US economy. The central case range provided is that average abatement cost in the power sector alone is \$45–61 per metric ton of CO_2. Even at the high abatement cost scenario in the US-REGEN results, the average abatement cost is still only \$36–87 per

metric ton of CO_2. These numbers are a big result, as the common standard is to compare them to the social cost of carbon dioxide. Based on the results of the best evidence available today, in Rennert and others (2022), the central estimate of the social cost of carbon dioxide is $185 per metric ton of CO_2 in 2020 (and would be slightly greater in 2030). The range in Rennert and others (2022) is from $120 to $400 per metric ton of CO_2 in 2030, with the variation here based on the discount rate.

Thus, these results indicate that the cost of abatement from the power sector provisions in the IRA is much less than the social cost of carbon dioxide emissions. In short, these abatement opportunities appear to be a great deal! There are numerous caveats laid out in the paper, but this basic finding speaks very well for the IRA. It also makes sense because renewable energy is becoming increasingly inexpensive and is already at a lower cost (on a levelized cost basis) than new fossil fuel power plants in most situations.

The authors notably do not calculate the cost-effectiveness of the many other tax credits in the IRA. There is a mention that the electric vehicle tax credits are likely to be less cost-effective due to a high degree of inframarginal customers who would have purchased electric vehicles anyway. But otherwise, the paper provides very little guidance on the cost-effectiveness of the rest of the bundle of tax credits. This is a glaring omission but understanding the abatement cost of the entire set of tax credits (or at least the largest ones) is of utmost importance for policy. One might speculate that some extremely cost-effective tax credits may be bundled with some much less cost-effective credits. As is discussed by Gillingham and Stock (2018), there is substantial heterogeneity in the abatement cost from different measures to reduce greenhouse gas emissions. The IRA has a smattering of different measures, including some that may not reduce greenhouse gas emissions at all, such as some of the hydrogen tax credits. Thus, very important work remains to assess the cost-effectiveness of all of the bundle of tax credits included in the IRA.

The impact of the IRA's clean power provisions. One striking finding by Bistline, Mehrotra, and Wolfram is displayed in figure 10 in the paper. This figure shows that under the IRA somewhere in the range of 22 to 45 percent of all hours over the year in the Southwest Power Pool region (primarily composed of South Dakota, Nebraska, Kansas, and Oklahoma) will have zero or negative wholesale prices of electricity. This comes about because so much renewable generation comes online and such renewable generation can operate at zero or negative prices due to the production tax credit. This is an incredibly striking result that occurs even with deployment of more

battery storage to store electricity from the worthless hours to use in the more valuable hours.

Having up to 44 percent of hours with zero or negative prices would make it difficult for baseline power plants, such as natural gas combined cycle plants, to stay online. Peaking plants could continue to be profitable due to the occasional times with high prices and low standby maintenance costs. But if so many hours become zero or negative priced, one would expect a dramatic reshuffling in electricity markets, possibly accompanied by the retirement of a large capacity of efficient baseload power. This issue is often known as the "missing money" problem (Hogan 2019) because the wholesale prices are so low so often that there just isn't enough revenue to keep baseload power plants online. Capacity markets in many regions pay power plants to remain active and may at least help prevent retirements—but potentially at a high cost.

What the authors do not do is tell us whether the issue observed in figure 10 is an extreme case or if it is prevalent throughout the electricity system. Even if it is an extreme case, it is a very important point. But if it is prevalent throughout the electricity system, this is a crucial insight for electricity system planners and policymakers. What it might mean is that capacity markets would have to provide substantial payments to certain key dispatchable or baseload power plants to keep them online (most likely fossil fuel plants). Alternatively, massive investments in energy storage would need to be made. This would substantially raise the overall total system cost of the electricity system. Presumably, US-REGEN already accounts for this in its detailed electricity system modeling. But a key insight from this consideration of grid integration and reliability costs is that retail electricity rates—which must pay for reliability—may not necessarily drop, even with the generous tax credits for renewables, which are effectively subsidizing electricity generation (and thus would be expected to reduce electricity rates).

Macroeconomic implications. Bistline, Mehrotra, and Wolfram provide a further contribution by exploring the macroeconomic implications of the IRA in two ways. First, they examine the macroeconomic impact of the IRA itself. In 2021, the US GDP was $23.32 trillion.[2] The total cumulative clean energy expenditure—while the largest outlay for clean energy in history—is going to be on the order of hundreds of billions of dollars over nearly a decade. Total spending per year will likely be less than $50 billion

2. FRED, Federal Reserve Bank of St. Louis, "Gross Domestic Product," https://fred.stlouisfed.org/graph/?g=15TKS.

at most. This is not pocket change, but relative to the size of the economy, it is quite small.

Thus, it may not be much of a surprise that the authors find minimal macroeconomic impacts. Of course, they perform their analysis in a much more rigorous fashion, using the Federal Reserve Board's US model (FRB/US), with inputs from US-REGEN. The model shows that the IRA is likely to increase demand initially and then later raise the capital stock and output. This result is sensible, but the magnitudes are so small that they are nearly a rounding error in macroeconomic aggregates.

The more important point that the authors make is that macroeconomic aggregates can substantially affect the implications of the IRA for clean energy investment. With a higher interest rate environment—something we are currently seeing as of the date of this writing—it will be more difficult to arrange financing for the clean energy investments of the IRA. This will dampen investment and reduce the emission reductions and fiscal cost of the IRA. Similarly, higher costs of labor and materials can dampen investment.

The authors perform a series of useful explorations of the sensitivity of the clean energy investments of the IRA to macroeconomic conditions. Figure 19 in the paper best shows the effects. I take away two findings from this figure. First, and perhaps most important, is that regardless of the costs of labor and materials (the sensitivity being explored in this figure), the power sector is projected to be substantially decarbonizing. However, higher costs of labor and materials could slow the growth of clean energy in the power sector. The authors do not perform similar calculations for the IRA tax credits outside of the power sector, and it would be useful to know if these two findings hold more broadly for the IRA investments.

Stepping back, it appears that while macroeconomic aggregates could be very slightly affected by the IRA and macroeconomic conditions could in turn affect the IRA, the bigger story is that the power sector is clearly on a path to decarbonize. Coal plants are quickly being retired (although some remain in the US-REGEN simulations as late as 2040, a result that I find to be unlikely) and the capacity to generate hundreds of gigawatts of renewable energy is being built (Fasching and Ray 2023).

Conceptual framework for understanding social welfare. Bistline, Mehrotra, and Wolfram develop a simple macroeconomic-style model to provide a conceptual framework for the welfare analysis of the IRA. This framework is applied to highlight the differences between the subsidies in the IRA and carbon taxes. This is well-trodden ground, and I'm not sure that a theory model was entirely necessary. The main point is that

an optimal carbon tax is going to be preferable from a welfare standpoint unless there are substantial deployment externalities. This is intuitive because the economy-wide carbon tax provides incentives to internalize the carbon dioxide externality and assures that the prices in the economy are the prices that maximize social welfare. By increasing prices based on the carbon content of goods, the carbon tax provides incentives for reducing carbon intensity through energy efficiency, fuel switching, and energy conservation. Now, in the electric power sector it may not be so obvious that electricity prices are too low because retail electricity rates are often far from the socially optimal level (Borenstein and Bushnell 2022). Indeed, sometimes retail electricity rates are already too high, in which case, some of the advantage of carbon pricing is eroded.

The authors' model is clear, and the treatment of deployment externalities is new to the literature to the best of my understanding. The deployment externalities in the authors' model are learning-by-doing spillovers, whereby the cost of the technology declines for everyone when a single decision-maker deploys more of the technology. The intuition for how learning-by-doing spillovers can lead to an under-deployment of new technologies relative to the social optimum follows from previous work (Gillingham and Sweeney 2010). Specifically, if a technology exhibits a substantial amount of learning spillovers, then the technology will be deployed too little by the private market, and policy that increases deployment (including tax credits like those in the IRA) can increase social welfare.

If such deployment externalities are sufficiently large, it is possible that the social welfare gains from a subsidy approach like the IRA can exceed the social welfare gains from a carbon pricing approach. The authors do not emphasize this, but really, the socially optimal strategy would be a carbon price to internalize the carbon externality along with a deployment subsidy to internalize the deployment externality. In this sense, the IRA might be thought of as only half of what is needed to really address climate change in an economically efficient manner.

A final nice insight in the model is the inclusion of bottlenecks. Bottlenecks are real-world frictions that can reduce investment in clean energy. The novelty of the approach is a new distinction between market bottlenecks (those that manifest as a higher price of investment when investment goes beyond a certain threshold) and nonmarket bottlenecks (a constraint on investment that does not have an impact on price). This distinction turns out not to be important for the total amount of investment, as long as the threshold is the same. However, the distinction can be very relevant for the fiscal costs of the investment tax credit, for the higher price of investment

could lead to larger federal expenditures from the tax credit, which is given as a percentage of the total investment. Indeed, at very high prices of capital due to bottlenecks, very little investment may be happening at high fiscal costs. This is a classic Laffer curve result from public finance that is intriguing to consider. It does not seem likely to be empirically relevant. However, some mix of market and nonmarket bottlenecks probably will occur with the substantial increase in clean energy investment spurred by the IRA. The large interconnection queues for renewable energy currently underway in many regions of the United States are just one example of a nonmarket queue that could slow the growth of clean energy, hampering the effects of the IRA (Penrod 2023).

FIVE KEY POINTS TO TAKE AWAY I conclude my remarks by summarizing five key takeaways from this insightful paper by Bistline, Mehrotra, and Wolfram, each of which lends itself to areas of future work to help us more deeply understand the IRA.

The first takeaway is that there is quite massive uncertainty in both the emission reductions from the IRA and the fiscal costs of the IRA. There is a tremendous upside in the IRA for emission reductions if the cost of renewables, electric vehicles, heat pumps, and other technologies continues to decline. But there are many reasons why the uncertainty bounds on future projections are extremely wide. The authors mention all of the reasons: technology uncertainty, market acceptance uncertainty, regulatory uncertainty, and macroeconomic conditions. One way to take the main result of the paper is that it points out just how far off of the CBO/JCT projections we could be in a decade.

The second takeaway is that the power sector tax credits in the IRA are remarkably cost-effective. Based on the estimated ranges in the paper, the abatement cost for the IRA power sector tax credits is well below the best estimates of the social cost of carbon dioxide that are available. This is not necessarily an obvious result, but it is a major result worth emphasizing. It also raises further questions, such as how long we can expect abatement costs to remain this low and what uncertainties they are most sensitive to.

The third takeaway is that the power sector tax credits can lead to some striking effects on the electricity system. Wholesale electricity prices could be zero or negative for up to 40 percent of hours of the year in some regions. The authors would have benefited from providing details on how many regions this is likely to be an issue for. Zero or negative wholesale prices would pose a huge problem to many legacy fossil fuel plants and would likely lead to a major impetus for stationary energy storage and to substantial increased retirements in the absence of additional payments

(either side payments or payments through capacity markets). What it also means is that the costs of reliability and grid integration may be sufficiently high such that it is quite possible that the IRA tax credits for clean electricity actually raise electricity rates, counter to what might be expected for a subsidy.

A fourth takeaway is that the overall welfare effects of the IRA really hinge on the success of the tax credits spurring innovation through increased deployment of clean energy. If the IRA really can lead to large learning-by-doing spillovers, it can substantially improve social welfare. But to be a market failure, one must have both substantial learning by doing with further deployment and spillovers across firms that lead firms outside the innovating firm to benefit from the innovation. Given the remarkable industrywide technological leaps that have been occurring over the past decade in both renewable energy and electric vehicles as the technologies have been deployed, I find it eminently plausible that such induced innovation is possible. But understanding this more deeply is certainly a topic for future research.

A final takeaway is that there is likely to be substantial heterogeneity in the abatement costs from the IRA, which is a bundle of many different tax credits. The authors allude to the possibility that the abatement costs would be higher for electric vehicles than renewable energy because so many electric vehicle buyers would be inframarginal (this of course may be countered by the large innovation potential in electric vehicle technology). But what about the energy efficiency and heat pump tax credits? What about the hydrogen tax credits? Carbon capture and sequestration tax credits? All of these unanswered questions underscore the vast uncertainty about the cost-effectiveness of the IRA. It is quite likely that many of these provisions will be extremely different in their cost-effectiveness. A full understanding of the economics of the IRA is not possible without estimates of the abatement cost of these measures. I see very high value in future work to begin peeling back the layers of the IRA to reveal which parts of the bundle are most likely to be cost-effective at reducing emissions.

Bistline, Mehrotra, and Wolfram break new ground in the depth and breadth of analysis on the IRA, which provides a historic expenditure to reduce greenhouse gas emissions. They uncover several striking findings but also leave open an important set of questions that are necessary to answer for a complete understanding of the IRA, which is crucial as the United States and the world consider the next steps in reducing greenhouse gas emissions.

REFERENCES FOR THE GILLINGHAM COMMENT

Borenstein, Severin, and James B. Bushnell. 2022. "Do Two Electricity Pricing Wrongs Make a Right? Cost Recovery, Externalities, and Efficiency." *American Economic Journal: Economic Policy* 14, no. 4: 80–110. doi.org/10.1257/pol.20190758.

Cole, Cassandra, Michael Droste, Christopher R. Knittel, Shanjun Li, and James H. Stock. 2023. "Policies for Electrifying the Light-Duty Vehicle Fleet in the United States." *American Economic Association Papers and Proceedings* 113:316–22. doi.org/10.1257/10.1257/pandp.20231063.

Congressional Budget Office (CBO). 2022. "Estimated Budgetary Effects of Public Law 117-169, to Provide for Reconciliation Pursuant to Title II of S. Con. Res. 14." Washington. www.cbo.gov/publication/58455.

Fasching, Elesia, and Suparna Ray. 2023. "Coal and Natural Gas Plants Will Account for 90% of U.S. Capacity Requirements in 2023." Washington: US Energy Information Administration. https://www.eia.gov/todayinenergy/detail.php?id=55439.

Forsythe, Connor R., Kenneth T. Gillingham, Jeremy J. Michalek, and Kate S. Whitefoot. 2023. "Technology Advancement Is Driving Electric Vehicle Adoption." *Proceedings of the National Academy of Sciences* 120, no. 23: e2219396120. doi.org/10.1073/pnas.2219396120.

Gillingham, Kenneth T., and James H. Stock. 2018. "The Cost of Reducing Greenhouse Gas Emissions." *Journal of Economic Perspectives* 32, no. 4: 53–72. doi.org/10.1257/jep.32.4.53.

Gillingham, Kenneth T., and James Sweeney. 2010. "Market Failure and the Structure of Externalities." In *Harnessing Renewable Energy in Electric Power Systems: Theory, Practice, Policy*, edited by Boaz Moselle, Jorge Padilla, and Richard Schmalensee. Washington: RFF Press.

Gillingham, Kenneth T., Arthur A. van Benthem, Stephanie Weber, Mohamed Ali Saafi, and Xin He. 2023. "Has Consumer Acceptance of Electric Vehicles Been Increasing? Evidence from Microdata on Every New Vehicle Sale in the United States." *American Economic Association Papers and Proceedings* 113:329–35. doi.org/10.1257/pandp.20231065.

Goldman Sachs. 2023. "Electric Vehicles Are Forecast to Be Half of Global Car Sales by 2035." February 10. www.goldmansachs.com/intelligence/pages/electric-vehicles-are-forecast-to-be-half-of-global-car-sales-by-2035.html.

Hogan, William W. 2019. "Market Design Practices: Which Ones Are Best? [In My View]." *IEEE Power and Energy Magazine* 17, no. 1: 100–104. doi.org/10.1109/MPE.2018.2871736.

IPCC. 2023. *AR6 Synthesis Report: Climate Change 2023*. Geneva. www.ipcc.ch/report/sixth-assessment-report-cycle/.

Jenkins, Jesse D., Erin N. Mayfield, Jamil Farbes, Ryan Jones, Neha Patankar, Qingyu Xu, and Greg Schivley. 2022. *Preliminary Report: The Climate and Energy Impacts of the Inflation Reduction Act of 2022*. Princeton, N.J.: REPEAT

Project. https://repeatproject.org/docs/REPEAT_IRA_Prelminary_Report_2022-08-04.pdf.

Lindsey, Rebecca, and LuAnn Dahlman. 2023. "Climate Change: Global Temperature." National Oceanic and Atmospheric Administration, January 18. www.climate.gov/news-features/understanding-climate/climate-change-global-temperature.

National Academies of Sciences, Engineering, and Medicine. 2021. *Assessment of Technologies for Improving Light-Duty Vehicle Fuel Economy—2025–2035*. Washington: National Academies Press. doi.org/10.17226/26092.

O'Boyle, Mike, Dan Esposito, and Michelle Solomon. 2022. *Implementing the Inflation Reduction Act: A Roadmap for State Electricity Policy*. San Francisco: Energy Innovation. https://energyinnovation.org/wp-content/uploads/2022/10/Implementing-the-Inflation-Reduction-Act-A-Roadmap-For-State-Policy-1.pdf.

Penrod, Emma. 2023. "US Grid Interconnection Backlog Jumps 40%, with Wait Times Expected to Grow as IRA Spurs More Renewables." Utility Dive, April 11. www.utilitydive.com/news/grid-interconnection-queue-berkeley-lab-lbnl-watt-coalition-wind-solar-renewables/647287/.

Rennert, Kevin, Frank Errickson, Brian C. Prest, Lisa Rennels, Richard G. Newell, William Pizer, and others. 2022. "Comprehensive Evidence Implies a Higher Social Cost of CO_2." *Nature* 610:687–92. doi.org/10.1038/s41586-022-05224-9.

Roy, Nicholas, Maya Domeshek, Dallas Burtraw, Karen Palmer, Kevin Rennert, Jhih-Shyang Shih, and Seth Villanueva. 2022. *Beyond Clean Energy: The Financial Incidence and Health Effects of the IRA*. Washington: Resources for the Future. https://media.rff.org/documents/Report_22-11_v5.pdf.

Schaffner, Carly. 2023. "Hyundai Leans on EV Tax Credit Lease Provision." *Automotive News*, April 14. https://www.autonews.com/retail/hyundai-uses-ev-tax-credit-leasing-provision-stay-competitive.

Senate Democrats. 2022. "Joint Statement from Leader Schumer and Senator Manchin Announcing Agreement to Add the Inflation Reduction Act of 2022 to the FY2022 Budget Reconciliation Bill and Vote in Senate Next Week." Washington. https://www.democrats.senate.gov/newsroom/press-releases/senate-majority-leader-chuck-schumer-d-ny-and-sen-joe-manchin-d-wv-on-wednesday-announced-that-they-have-struck-a-long-awaited-deal-on-legislation-that-aims-to-reform-the-tax-code-fight-climate-change-and-cut-health-care-costs.

GENERAL DISCUSSION James Stock began the discussion by noting that there is additional support for the authors' findings in the US Energy Information Administration's annual energy outlook for 2023, which estimated that the Inflation Reduction Act (IRA) would have a 7 percentage point marginal effect on emissions by 2030.[1]

1. US Energy Information Administration, *Annual Energy Outlook 2023: Issues in Focus—Inflation Reduction Act Cases in the AEO2023* (Washington: Department of Energy, 2023), https://www.eia.gov/outlooks/aeo/IIF_IRA/pdf/IRA_IIF.pdf.

Kenneth Rogoff discussed estimates showing that even with the IRA, the United States is not making sufficiently significant emission reductions to have a significant global impact and that the overwhelming growth in future emissions is coming from elsewhere.[2] He acknowledged the importance of leading by example but questioned the term "cost-effective" given that the relative impact on global emissions was not discussed in greater depth. Rogoff stated that he found the learning-by-doing aspects of the IRA powerful and important in a global context. Maurice Obstfeld agreed that while US efforts are small relative to the challenge, doing what is politically possible is still important. Obstfeld also noted that one potentially positive impact of this legislation is that the European Union could make green subsidies more accessible to remain competitive with green investment in the United States.[3]

Steven Davis asked whether the authors' carbon abatement cost estimate was for within the United States or an estimate of how it would affect global emissions abatement. He expressed concerns over the export, and potentially lower prices, of fossil fuels abroad if the United States reduces its reliance on carbon energy sources. Davis agreed with Rogoff's questioning the use of national measures of impact when many of the outcomes of interest are global. Davis also supported the benefits brought up by Rogoff and Obstfeld, stating that induced innovation and international policy bandwagon effects were potentially important positive aspects of the legislation. Catherine Wolfram responded to Davis's concerns about exporting emissions abroad, stating that estimates of leakage are low, particularly at a price of $12 per ton of carbon.[4]

Glenn Rudebusch recognized that while the macroeconomic benefits of less carbon pollution may appear to be small, there are important co-benefits of less carbon pollution for health and labor force participation in addition to distributional effects, based on where carbon pollution is occurring. Rudebusch also mentioned the importance of directed technical change, suggesting that the authors might include greater discussion of innovation externalities. Pinelopi Goldberg highlighted the importance of the IRA and noted that until recently the United States was one of the only countries

2. Climate Action Tracker, "USA," https://climateactiontracker.org/countries/usa/.

3. Jacob Funk Kirkegaard, "The US-EU Race for Green Subsidies Can Help Fight Climate Change," Realtime Economics, Peterson Institute for International Economics, February 14, 2023, https://www.piie.com/blogs/realtime-economics/us-eu-race-green-subsidies-can-help-fight-climate-change.

4. Florian Misch and Philippe Wingender, "Revisiting Carbon Leakage," working paper 21/207 (Washington: International Monetary Fund, 2021).

that had not made significant progress toward addressing climate change—despite being the second-largest carbon emitter and one of the richest countries in the world.[5]

Stock responded to the comments made by Rogoff and Davis, stating that he doubted they would have asked, "Why should we adopt a cost beneficial policy in the United States to reduce emissions when most of the emissions are abroad?" if instead the question was, "Why should we adopt a cost-beneficial policy to help the uneducated poor children of the United States when most uneducated poor children are abroad?" Stock continued, explaining that four reasons to adopt such a policy are: (1) it is cost beneficial, (2) the political economy is such that those experiencing the impacts abroad do not vote in the United States, (3) there will be positive spillover effects from involvement in international negotiations such as the Paris agreement, and (4) it will help advance technological innovations through learning by doing.

Wolfram responded to comments regarding spillover effects, noting the importance of learning by doing as well as spillovers with pricing, such as the carbon border adjustment mechanism implemented in the European Union, which has encouraged other countries to adopt carbon pricing. John Bistline noted that prior literature examining the impacts of induced technical change in wind and solar will likely be important for thinking about the impacts—via mechanisms such as learning by doing, economies of scale, and research and development subsidies—of the IRA going forward.

John Haltiwanger mentioned the importance and need to further examine the distributional consequences and sectoral shifts mentioned by Jason Furman in his comment, particularly the between-sector effects and within-sector effects. Haltiwanger wondered if subsidies will mostly go to large incumbent businesses, which could have adverse effects for innovation and productivity growth that are worth recognizing and discussing. Adele Morris emphasized the importance of Furman's discussion of distributional effects and the potential for subsidies, such as those for electric vehicles (EVs), to be potentially regressive, complicating some of the benefits of learning by doing.

Wolfram stated that the distributional effects highlighted by Furman and Kenneth Gillingham are very important and pointed out that this work is a first analysis of the IRA and that there will likely be many papers written

5. Monica Crippa, D. Guizzardi, M. Banja, E. Solazzo, M. Muntean, E. Schaaf, and others, *CO2 Emissions of All World Countries* (European Commission, 2022), https://edgar.jrc.ec.europa.eu/report_2022.

on the IRA discussing aspects of the legislation going forward. Bistline responded to these remarks as well, highlighting previous work that had examined how electrification incentives could have distributional impacts for local air quality.[6]

Phillip Swagel pointed out that legislation that generates changes in the Internal Revenue code are estimated by the Joint Committee on Taxation (JCT), not the Congressional Budget Office (CBO).[7] Swagel stated that creating preliminary estimates takes time and is challenging, particularly in the case of the IRA, which was a prolonged legislative effort. Swagel also responded to Gillingham's discussion of implementation regarding rules of origin and tax credits only applying to certain limits, noting that this may not be a fair critique of the JCT. He speculated that the JCT may have evaluated the rules as the statute stipulated but that the Treasury instead determined that income limits do not apply under specific circumstances.

Morris expressed the need to distinguish between fiscal cost and real resource cost, particularly when discussing the cost per ton abated. She remarked that in some instances abatement cost is the combined private and fiscal cost on average whereas a carbon tax sets the marginal abatement cost of an incremental environmental improvement. Morris also affirmed examining the legislation in the broader regulatory framework, noting that if EV makers can sell credits into the corporate average fuel economy credit market, this could have a significant impact on the net environmental benefits of an EV subsidy, relative to a counterfactual where EV makers cannot sell these credits.

Joe Beaulieu explained that he was surprised by the expected wholesale prices becoming negative and was reminded of the 2000–2001 California electricity crisis in which abrupt changes in the energy market structure had unexpected results.[8] Beaulieu wondered if changes to the market structure of the energy system from the IRA might be relevant to discuss or consider in this work. Wolfram responded to Beaulieu by clarifying that while the authors predict negative wholesale prices, the figure shown was for one region with a lot of wind capacity and that this does not reflect

6. John E. T. Bistline and others, "Economy-Wide Evaluation of CO2 and Air Quality Impacts of Electrification in the United States," *Nature Communications* 13 (2022): 6693, https://www.nature.com/articles/s41467-022-33902-9.

7. Joint Committee on Taxation, "Statutory Basis," https://www.jct.gov/about-us/statutory-basis/.

8. James L. Sweeney, "The California Electricity Crisis: Lessons for the Future," *The Bridge* 32, no. 2 (2002): 23–31, https://www.nae.edu/Publications/Bridge/OurEnergyFuture/TheCaliforniaElectricityCrisisLessonsfortheFuture.aspx.

prices nationwide. Tristan Reed pointed out that large increases in capital investment imply an increase in demand for construction which, in an environment with low unemployment, could lead to a bottleneck in implementation. Reed contemplated whether this was something the authors could explore or discuss in their model.

Goldberg noted that while it is very useful to compare policies to a carbon tax, carbon taxes have not been implemented in the past. Goldberg encouraged a greater discussion of alternative policies to a carbon tax that might be more politically feasible. Wolfram responded to Goldberg by pointing out that a number of countries, including the European Union, United Kingdom, and Canada have implemented carbon prices, in addition to states in the United States such as California and others in the Northeast.[9] Stan Veuger commented that he believed that the motivation for the IRA was not to increase innovation externalities but instead a result of political constraints, such as a lack of support for carbon taxes as well as protectionist goals of the government. Veuger explained that he believed this was worth highlighting because of the international response to the legislation.

Wolfram responded to Furman's comment by adding some caution to the work of Bornstein and Kellogg, stating that their finding that the retail price of electricity had been higher than the social marginal cost was based on old social cost of carbon numbers.[10] Wolfram also noted that it may not be an equal comparison to scale the CBO/JCT estimates to the authors' estimates due to the difference in the EV tax credits.

9. World Bank, "Carbon Pricing Dashboard," https://carbonpricingdashboard.worldbank.org/map_data.

10. Severin Borenstein and Ryan Kellogg, "Carbon Pricing, Clean Electricity Standards, and Clean Electricity Subsidies on the Path to Zero Emissions," *Environmental and Energy Policy and the Economy* 4 (2023): 125–76.

PANEL on EMERGING MARKET CHALLENGES

Note: Papers prepared for this panel were guest-edited by Maurice Obstfeld.

VIRAL V. ACHARYA
NYU Stern School of Business

India at 75: Replete with Contradictions, Brimming with Opportunities, Saddled with Challenges

ABSTRACT I present a perspective on where the Indian economy stands right now. I acknowledge the contradictions that have arisen given the divergent growth path of urban, formal or (stock-market) listed India relative to rural, informal or unlisted India. I also focus on the country's immense opportunities in expanding the digital footprint of finance to last-mile borrowers. I present novel facts on the rising industrial concentration, drawing out its historical evolution, the channels that have caused it to rise recently, and its implications for product price markups and inflation. I recommend that to restore industrial balance, India increase overall competition by reducing import tariffs and reduce the pricing power of its largest conglomerates. I also propose that to restore macroeconomic balance, India reduce fiscal deficit and public sector borrowing requirements as well as rein in inflation, address gaps in skills and education, and restore female labor force participation.

I. Replete with Contradictions

India is complex and can be hard to fathom. To illustrate this enduring fact, I start by presenting four contradictions—or just counterintuitive or not so obvious juxtapositions of phenomena—currently at play in the Indian economy.

Conflict of Interest Disclosure: Viral V. Acharya was deputy governor of the Reserve Bank of India from January 2017 to July 2019. 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. The author is not currently an officer, director, or board member of any organization with a financial or political interest in this paper.

Brookings Papers on Economic Activity, Spring 2023: 185–215 © 2024 The Brookings Institution.

I.A. Stock Market versus Real Economic Performance

Perhaps the most salient contradiction around India concerns its staggering post-COVID-19 stock price run-up relative to the strength of its economic recovery, both in an absolute sense as well as relative to other emerging market (EM) economies. For instance, the large plus mid-cap Indian stock market (MSCI World) index has risen since January and April 2020 by 75 percent and 100 percent respectively (until March 2023), whereas the corresponding MSCI EM ASEAN index has been flat and risen by 50 percent only (see figure 1 in the online appendix).

This *financial* outperformance stands in stark contrast to India's *real* economic underperformance since the pandemic: its GDP level is still 6 percent below that implied by the pre-pandemic trend applied to January 2020 GDP, and the employment-to-population ratio has stayed just above 50 percent in the last three years as per data from the Periodic Labor Force Survey (PLFS) but has fallen from around 40 percent prior to the pandemic to 37 percent as per data from the Centre for Monitoring Indian Economy (Chinoy, Jain, and Sood 2022).[1] The private consumption path remains below the pre-COVID-19 potential path as per data from the Ministry of Statistics and Programme Implementation (MoSPI), and household spending plans—though improving—still remain below pre-COVID-19 levels in the Reserve Bank of India's (RBI) consumer confidence surveys (Chinoy, Jain, and Sood 2022).[2]

This scarring, even three years after the pandemic, has occurred in spite of the services exports boom in the post-pandemic recovery and pent-up consumer demand having now played out fully. The weak economic performance reflects the so-called *K*-shaped nature of the recovery wherein urban and formal India recovered particularly well but rural and informal India lagged behind (as explained below). It likely also reveals pre-pandemic weaknesses.

Contrasting this, however, is the fact that India is projected to be a bright spot in the world economy going forward (together with China). For example,

1. India's GDP (current prices) in 2023 is US$3.74 trillion; International Monetary Fund, "GDP, Current Prices," https://www.imf.org/external/datamapper/NGDPD@WEO/GBR/IND. In 2020, the GDP (current US$) was US$2.67 trillion. In 2019, India's GDP was US$2.8 trillion; World Bank, "GDP (Current US$)—India," https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2021&locations=IN&start=2019. See also Centre for Monitoring Indian Economy Pvt. Ltd., https://www.cmie.com/.

2. https://www.mospi.gov.in/; https://m.rbi.org.in/scripts/BimonthlyPublications.aspx?head=Consumer%20Confidence%20Survey%20-%20Bi-monthly.

the International Monetary Fund's projections in April 2023 for the year 2023 and 2024 for India are respectively 5.9 percent and 6.3 percent (5.2 percent and 4.5 percent for China).[3] Importantly, even if on a lower post-pandemic base as explained above, India is projected to accelerate, while advanced economies (and even China) decelerate.

I.B. Urban versus Rural and Formal versus Informal India

Depending on which part of India one talks about, it is either booming (urban and formal sector) or it still remains scarred from the pandemic (rural and informal sector). Since the pandemic, the operating profit-to-sales gap has opened up widely between large firms (+1.5 percent) versus small firms (−0.5 percent), the latter being defined as having less than 50 million Indian rupees (INR) or $60,000 in capital (Bhandari and Chaudhary 2022). Besides the direct effect of lockdowns, smaller firms have been hit thereafter by rising commodity prices having an adverse impact on their (large) share of raw material to sales.[4] As a result, while large manufacturing firms have been able to retain their size, small manufacturing firms have contracted by 14 percent (Bhandari and Chaudhary 2022). Correspondingly, the performance of the listed large companies has been stellar, and they have been able to grow significantly, in part at the expense of smaller firms, further explaining the lack of congruence in financial versus real performance of the Indian economy.

Why does this divergence matter? Small firms and establishments are important in India for employment. In fact, as is the case in many other countries, like the United States, they contribute to over 40 percent of overall labor in India. Separately, 40 percent of labor in India is also in agriculture (Bhandari and Chaudhary 2022). Both small firms and agriculture tend to have a greater presence in rural India than in urban India. Hence, the divergent performance of large and small firms has immediate implications for urban and rural demand. In particular, weak rural demand is seen notably in (1) rural unemployment insurance demand (demand for work under the Mahatma Gandhi National Rural Employment Guarantee Act) doubling at the onset of the pandemic and remaining elevated for more than two years thereafter before gradually normalizing to pre-pandemic levels (Chinoy, Jain, and Sood 2022), and (2) sales of two-wheelers significantly

3. International Monetary Fund, "Real GDP Growth," https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/FLERRC/FXERRC/CHN/IND.

4. See, for instance, Mohommad and others (2023).

underperforming those of passenger vehicles and ultra-premium bikes.[5] In turn, while wages in urban India remain elevated far in excess of the inflation rate, rural wages, even if increasing nominally, have been outpaced by inflation for most of 2022 (Chakraborty and Baqar 2023). This fall in real rural wages has interacted with shocks to commodity prices and weather to create a weak rural, informal, small-firm economy.

I.C. Goods Inflation in India

Bhandari and Chaudhary (2022) also point out that demand for goods in India has remained weak after the pandemic, even as services demand has grown; however, goods prices in India have remained elevated, even after global goods inflation softened in 2022 as supply chain issues eased. Indeed, goods inflation momentum in India is positive whereas its global counterpart is negative. What explains this lack of congruence between demand and prices, and between India and the rest of the world? It is argued that goods inflation in India is in fact likely to persist as margins of manufacturers in India are substantially high due to their protection via tariffs (Chatterjee and Subramanian 2020) and their market power from rising industrial concentration (Chakraborty and Baqar 2023). In contrast, tariffs and margins in services sectors are smaller, even though there are emerging signs of concentration in some services sectors too, such as in telecommunications.

I.D. Fewer Defaults, but Under-Recovering, Bankruptcy Resolution Process

Depending on who you speak to, the Insolvency and Bankruptcy Code (IBC), enacted in 2016 and operational since 2017, has been yet another failure in corporate insolvency resolution or a resounding success.

The critical view of the IBC stems from the facts that, first, debt in India continues to perform closer to equity as lamented by Vishwanathan (2018), recovering only 39 percent for resolutions under the IBC (Gupta and others 2021). The recovery rate comes down to only 24 percent if the largest nine IBC cases are excluded. This makes recoveries for bank loans in India virtually half of the global average. Second, average time from filing of a case to resolution has been 561 days, about twice what was originally envisaged under the IBC. And third, a phenomenal 45 percent of the cases under the

5. Society of Indian Automobile Manufacturers, "Performance of Auto Industry in 2022–23," https://www.siam.in/statistics.aspx?mpgid=8&pgidtrail=9.

IBC get liquidated. Further, several cases are filed without resolution plans and liquidations recover only 7 percent for creditors. Unsurprisingly, several aspects of IBC remain under continued legal scrutiny and revision.[6]

The salubrious view of the IBC arises from the observations that the Indian banking sector has now brought down its nonperforming assets ratio significantly by resolving cases under the IBC, recognizing losses, recapitalizing balance sheets, and now being ready to provide credit for growth at healthy rates and quantities (see II.E). As a result of the banking sector cleanup, capacity utilization of distressed sectors has improved and overall risen to close to 75 percent at present from a low of 60 percent prior to 2017, when there was an oversupply of zombie firms in these sectors (see figure 2 in the online appendix).[7] Fresh slippages into nonperforming loans have declined due to an important deterrence effect of the IBC and decisive regulatory actions by the Reserve Bank of India (RBI), whereby loss of control for business owners and individual promoters has led to a de-leveraging of the Indian corporate sector. The debt-to-GDP ratio of the Indian corporate sector has declined over a decade from 78 percent to 50 percent (Chinoy, Jain, and Sood 2022).

In sum, while creditor recoveries have not been much healthier under the IBC compared to prior bankruptcy codes in India, there has been a meaningful loss of control for asset owners and this has facilitated a healthier credit economy. Grievances around IBC nevertheless abound.

II. Brimming with Opportunities

Notwithstanding these contradictions or puzzling facts, there is an unmistakable entrepreneurial spirit feverishly at work in India. It has taken hold over the past decade on the back of India's digital plumbing and equity market deepening. Furthermore, the restoration of banking (and nonbanking) sector health to adequate capital standards, along with the advances in digital and fintech lending, augurs particularly well for credit to small firms.

6. The pandemic itself raised difficulties around the IBC resolution path for investors. The government suspended any fresh applications to the IBC for a year after the March 25, 2020, lockdown. Further, the RBI, along with a slew of rate cuts and effective liquidity measures, also introduced debt moratoria on payment on term loans and deferment of interest on working capital (Mohan 2021).

7. Kulkarni and others (2019) report that the percentage of zombie firms in the RBI's CRILC (Central Repository of Information on Large Credits) bank-borrower credit data was 21.6 percent during March 2016 to March 2019.

II.A. Start-up India

Top candidates from the Indian Institutes of Management or Technology are no longer keen to do a PhD in finance or economics. It is more likely that they want to be entrepreneurs and start something of their own, typically related to information technology (IT) services. India is busy churning out unicorns (start-ups with market valuations above $1 billion) aplenty. This spell has been steady over the past two decades but has accelerated since the foundations of "Digital India" have been put in place over the past decade, and its canvas has gradually become diverse across a range of sectors but predominantly within services. India now ranks fourth behind the United States, China, and Europe when it comes to the number of unicorns (Mishra and others 2021a, 2021b). Over time, market capitalization and the proportion of unicorns listed on the stock market has grown.

When the first generation of modern Indian entrepreneurs was born following the liberalization of the 1990s, it featured companies that primarily customized, installed, managed, and maintained software such as SAP for the rest of the world or that engaged in outsourcing, taking over the world's back-office operations, telemarketing, or customer services. Out of this original set grew the giants of today, such as Infosys, Tata Consultancy Services, Wipro, and Tech Mahindra. Once this set of entrepreneurs and their employees figured out what the rest of the world wanted, they grew in confidence and acquired software development capabilities which led to the second generation of entrepreneurs in India. This generation designs programs and solves problems for the rest of the world. However, there is now a third generation of entrepreneurs in India that is catering not to the rest of the world but to the Indian consumer. This set of companies is focused on e-commerce, fintech, edtech, and all forms of digital services, and has exploded since the pandemic, on the back of India Stack. Given the size of the Indian consumer base, the sky appears to be the limit for this third generation, at least for the near future.

II.B. Digital India

India, by most objective standards, has the world's best digital plumbing, which has evolved modularly. It is designed around Aadhaar, the unique identity number rolled out starting in September 2010, which has now been provided to over 1.2 billion Indians with several scores of millions of authentications occurring daily (Totty 2022). This has enabled, after early digital payment mechanisms, the setting up of the Unified Payments Interface (UPI), a public utility that provides a common payments

and settlements platform between any two Indian entities (individuals or businesses) with unique identifiers, enabling seamless portability between their respective banking solutions at the back end. Combining this with the deep penetration of mobile technology in India has then led to the delivery of digital payments–based solutions, e-commerce, e-KYC (Know Your Customer), e-signing of documents, and the like. This public-utility approach adopted by India has differed from the mostly private but concentrated model of digitization in China and the private but heavily fragmented and seemingly inefficient model of digitization in the United States.

The net result is that digital payments in India are rising in share at the expense of cash, which since 2020 is declining in its overall share of payments for the first time in India, representing a substantial turning point for the economy (see figure 5 in the online appendix). While the ill-conceived demonetization of 2016 failed to create a definitive transformation from cash to digital payments, the well-intended even if botched rollout of the goods and services tax (GST) has moved even smaller businesses onto digital platforms such as GST Network and the Government e-Marketplace (GeM). Post pandemic, however, the transformation has finally touched Indian households decisively. There are now more than two hundred million active users of digital payments, with Google Pay and PhonePe making over 80 percent of the transactions (Mishra and others 2021a, 2021b). What is even more impressive is that India continues to push ahead along this guiding principle of viewing payments- and settlements-related services as a public good that many observers contend indeed ought to be provided publicly. India Stack, in particular, provides a set of application programming interfaces (APIs) that are standardized to ensure encrypted trustworthy pipes connecting various first-order and higher-order platforms for customer or business interfaces, developed with the support of India's tech think-tank iSPIRT (Indian Software Product Industry Round Table). Thinkers who have made Digital India such a success deserve an Olympic gold medal, even if digitally delivered!

II.C. FinTech in MSME Lending

Historically, India's micro, small, and medium-sized enterprises (MSMEs) have struggled to receive formal financing.[8] While the financial system has been creative with microfinance to improve access to credit, the report

8. See, for example, the 2019 UK Sinha Committee report of the RBI.

assessed that there remains a formal financing gap in MSME credit of over $3.5 trillion. It seems, however, that the situation is now changing on the ground.

First, from only 44 percent of Indian male citizens and 26 percent of Indian female citizens having a bank account in 2011, the proportion in 2021 grew to 78 percent each.[9] Second, there is now a data layer being added to India Stack, and the access to an "account aggregator" enables an entity to pool a digital view of all its financial holdings data, based on India Stack's secure consent architecture, for enabling algorithmic credit scoring. Third, MSME transactions are now captured electronically on private or government e-commerce platforms, making them readily collateralizable for account receivables financing. This has helped alleviate their liquidity risks. Finally, as India Stack has created an open credit enablement network (OCEN) that provides portable pipes between banks, sharing economy platforms, and end borrowers, entry barriers are low, and payment companies that are springing up at a fast rate are keen to evolve into fintech lenders to individuals and MSMEs by joining forces with e-commerce platforms that are eager to provide loan assistance.

This overall formalization of the MSMEs and the financial lending technology, built around India's rich and robust digital plumbing architecture, has implied that (1) credit access for MSMEs has substantially eased, consistent with new business and income growth in Indian districts where the adoption of cashless payments has been more intense since 2016 (Dubey and Purnanandam 2023); (2) fintech lending share in credit has grown with the fintech sector valuation at over $20 billion (Gupta and Shah 2021); and (3) private equity and venture capital funding to this sector is over $10 billion and represents the second-largest investment in the economy after the e-commerce sector (Gupta and Shah 2021). These developments have raised the possibility of reaching the last mile on the road to providing banking services for the 60 million plus MSMEs and more than a billion citizens of India over the next decade.

II.D. Financial Backbone of Start-up and Digital India

What attracts the private equity and venture capital, as well as angels—foreign and domestic—that invest in India's entrepreneurial economy? Besides the potential of the young Digital India firms, it is the relatively

9. Statista, "Gender-Wise Distribution of Bank Accounts in India from 2011 to 2021," https://www.statista.com/statistics/942795/india-financial-institution-account-ownership-rate/#:~:text=In%202021%2C%20about%2078%20percent,less%20educated%20and%20the%20poor.

vibrant, easy-to-exit primary and secondary market for equities. The development and deepening of equity markets has been aided by the relatively high, even if gradually declining, household savings rate and the financialization of savings away from real estate and gold.[10] This financialization occurred at a rapid pace over the past decade through the advent of mutual fund schemes and their penetration in second- and third-tier cities in India, and was induced to an extent by the tepid pace of the rise in real estate valuation as well as by the decline in inflation relative to the prior decade.

Reflecting this, India remains a large recipient of foreign direct investment (FDI) within emerging market countries, in absolute flows next only to China and Brazil, even though there has been some decline of late (Chinoy, Jain, and Sood 2022). Clearly, relaxing some more sectors for FDI is a natural way to ensure further penetration of FDI and the financing and value-add that these investors bring to the economy. Even at current levels, however, FDI has joined remittances in creating a stable inflow of foreign capital relative to the volatile foreign portfolio investment flows. A further potential stems from the global pivot toward the China Plus One strategy, which could bode well for FDI in manufacturing in India (Apple and Foxconn shifting part of their manufacturing of phones to India to diversify supply chains, for instance), but it remains to be seen if this potential will be fully realized or not.

II.E. Health of the Banking System

Finally, capitalizing on the IBC, a range of initiatives has been undertaken by the RBI since 2017 to resolve the nonperforming assets of the Indian banking system.[11] The efforts have now come to fruition in that gross and net nonperforming assets of scheduled commercial banks were down in September 2022 to 5 percent and 1.3 percent, respectively, having fallen from peaks of 11.5 percent and approximately 6 percent in 2018 (RBI 2018, 2022). Importantly, the provision coverage ratio for Indian banks now exceeds 70 percent, and fresh slippage ratios remain low given the deterrence effect of the IBC (see section I.D) and the de-leveraging of most large and listed companies (RBI 2022). It has taken more than ten years for this cleanup to materialize following the credit boom and bust of

10. The hoarding of gold by Indian households, mostly in the form of jewelry that serves the purpose of inheritance transfers and wedding gifts, dates back to ancient times. It was only entrenched by high inflation until the last decade's disinflation succeeded in inducing an aggregate switch toward financialization of savings. Nevertheless, the Indian demand for gold remains sensitive to fluctuations in inflation.

11. See Acharya (2020) for background.

2011–2013, especially in sectors such as infrastructure, power, ports, and steel. The good news from all this is that if private capital expenditures in India were to pick up, banks are in a position to meet the credit demand. The bad news is that this hasn't yet occurred.

III. Saddled with Challenges

A key question is, given India's opportunities are vast, will it be able to register higher growth and output levels consistent with its potential and expectations, create jobs at a pace and of a quality that its growing population requires, and become a greater part of the global economy? I wish to highlight several structural—industrial and macroeconomic—issues that India remains saddled with and which present daunting but surmountable challenges for the future.

III.A. Tariffs

India is undoubtedly a contender for being the "tariff king" of the world. As per World Trade Organization (WTO) records, India's average present tariff rate of greater than 15 percent (18.3 percent in 2021) is the fourth-highest behind Sudan, Egypt, and Venezuela, on par with Brazil, and substantially higher than China and Mexico (Chan 2019; Sinha 2023).[12] While India's tariff rate has no doubt come down from being above 50 percent prior to 1991, it has had no substantial decline since the global financial crisis of 2007–2009, and has in fact increased by about 5 percent since 2013.[13] As Aiyar (2018) notes, the present Indian government which came to power in 2014 with the slogan "Minimum government, maximum governance," has, however, reversed the liberalizing trend of its earlier term during 1998 to 2004. It has instead adopted protectionism, for example, via its budget in February 2018 which raised import duties significantly and across the board "in order to protect uncompetitive small businesses and create jobs in labor-intensive industries" (Aiyar 2018, 1).[14] Chatterjee

12. The World Bank, "Tariff Rate, Most Favored Nation, Simple Mean, All Products (%)," https://data.worldbank.org/indicator/TM.TAX.MRCH.SM.FN.ZS.

13. Macrotrends, "India Tariff Rates 1990–2023," https://www.macrotrends.net/countries/IND/india/tariff-rates.

14. Aiyar (2018, 1) notes: "Modi's Bharatiya Janata Party (BJP) is not a conventional right-wing party. It rejects both socialism and Western capitalism and seeks a homegrown solution called Integral Humanism. It supports private enterprise but also runs India's biggest trade union and believes in a wide-ranging welfare state. It has highly protectionist affiliates that have always been wary of multinational corporations and international institutions. It believes in government intervention to create national champions, increase employment, and protect small businesses."

and Subramanian (2020) document the time-series and sectoral patterns of India's tariffs and argue that this has instead hurt India's exports in labor-intensive sectors such as apparel, textiles, leather, and footwear, where India has ceded much ground to its neighbors. In other words, India is protectionist in precisely those sectors, namely, goods manufacturing, where the global opportunity has arisen from the Chinese slowdown and the China Plus One pivot of the global economy.

There are several harmful consequences. First, while India has become more self-reliant on agricultural output, tariffs in this sector remain close to 35 percent (Chatterjee and Subramanian 2020). At the same time, its efficiency remains low. For employing more than 40 percent of India's work force, agriculture generates approximately 16.8 percent of the GDP.[15] This prevents a market-based rotation of jobs in India from low-skilled agricultural labor to high-skilled services labor. Second, India exports to the rest of the world by processing and adding value to imported raw materials and goods. As a result, high tariffs—by increasing the cost of imports—have made exported goods by Indian firms costly and globally uncompetitive, lowering India's goods exports and in turn its share in global goods trade. It is hard, therefore, to find many products outside India that are manufactured by Indian firms. There are few, if any, global Indian brands. Third, high tariffs imply that Indians pay much more for many imported items (such as iPhones) than foreign consumers do, and in many cases have to simply put up with weaker-quality, higher-priced domestic substitutes. In turn, price levels in the economy are kept artificially high in spite of global efficiency gains that could aid disinflation. Finally, tariffs have created protectionism in several Indian industries, disincentivizing investments in efficiency by cozy incumbents and allowing them to steadily garner market power by building up concentrated positions.[16] There are, however, other factors besides tariffs that have contributed to this market power, for

15. The World Bank, "Agriculture, Forestry, and Fishing, Value Added (% of GDP)—India," https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?locations=IN; "Employment in Agriculture (% of Total Employment) (Modeled ILO Estimate)," https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS.

16. The political economy of why the declining trend in tariffs has reversed is interesting to explore. It is consistent with the populist messaging of making India self-reliant (*atmanirbhar* in Hindi), while also convenient for large domestic incumbent firms, who remain protected from foreign competition. It may, however, also be reflective of an inability to create adequate jobs given China's dominant role in goods exports and the threat of automation. Affected sectors, especially their MSMEs, have been regularly offered forbearance on their bank credit, so the real economic cost appears to be borne by the society at large via inflation, low presence in global trade, and a lack of investment and jobs. These generalized costs are difficult to pinpoint specifically to tariffs.

which I present next—to the best of my knowledge—a novel set of facts and analysis.

III.B. Concentration of Power in Indian Industry

India was effectively a closed economy until 1991 and industrial concentration was high due to state-owned monopolies.[17] Post-1991 liberalization had a dramatic impact on concentration as industries were de-reserved for the private sector and public sector enterprises were privatized or divested. As a result, industrial concentration, measured using the share of the top five industrial firms or groups across the nonfinancial sector by the highest sales or assets *in a given year*, fell dramatically to start with (see figure 1, panel A, for the share of the top five by assets, and the online appendix for all results based on share by sales). Essentially, public sector firms gave up their share to private entrants.

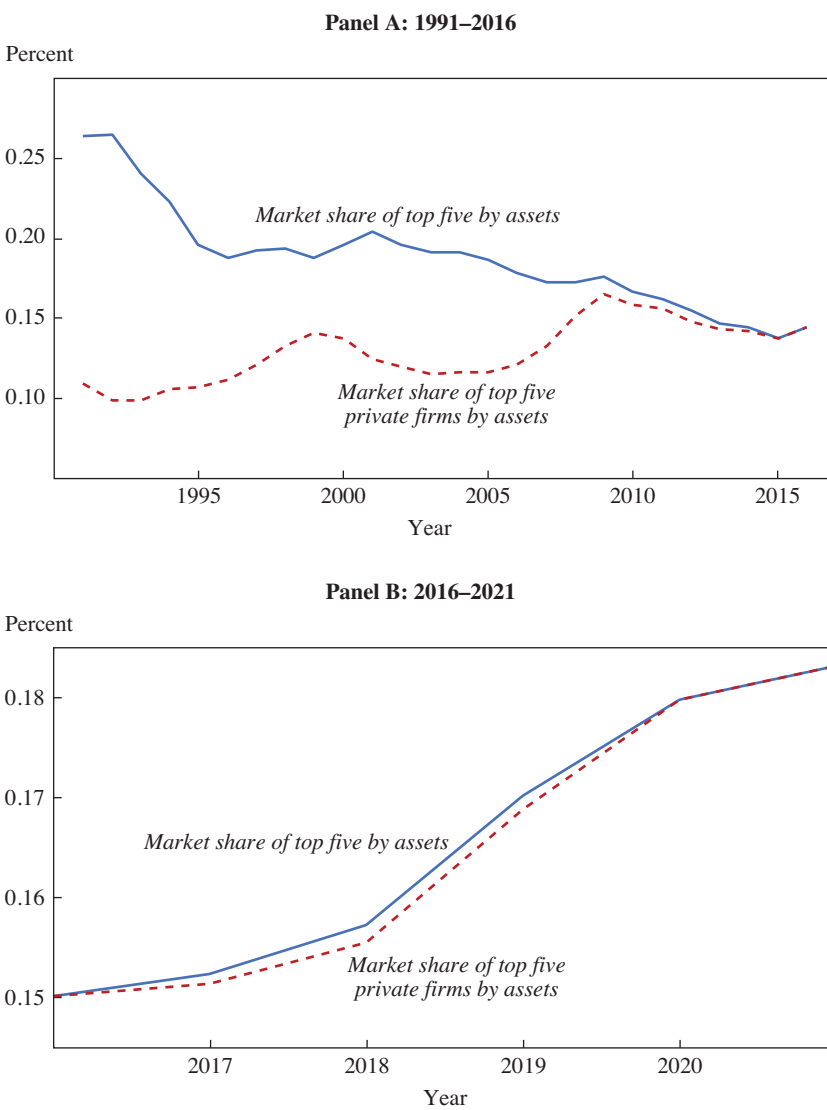
In other words, overall concentration fell, and even though concentration within the top five private firms gradually rose, it was low to start with. By 2010, the concentration of market power in the top five private firms had caught up with the concentration of overall top five firms (including public sector firms), but both fell during 2010–2015. Next, a close inspection shows that concentration started rising again from 2015 onward, overall as well as just within the set of private top five firms (figure 1, panel B). Put another way, private top five groups evolved into the overall top five across many nonfinancial sectors. At a disaggregated sectoral level, too, the notable shift occurs around 2015–2016 in several sectors, mostly traditional or capital-intensive (e.g., civil engineering; metals; nonmetal minerals; chemical, petroleum, and wood products and retail trade), but recently, also in newer sectors such as telecommunications.[18]

A striking feature of this rise in industrial concentration by private companies is that it has been in part due to the growing footprint of the "Big-5" industrial conglomerates, based on the highest overall share of assets in

17. The analysis that follows on India's industrial concentration is based on joint work with Rahul Singh Chauhan of the University of Chicago and uses the Prowess Dx database at the Centre for Monitoring Indian Economy (CMIE).

18. By way of specific examples, the share of the top five groups by sales in civil engineering and construction rose from 31 percent in 2016 to 42 percent in 2021, in telecommunications from 65 percent to over 84 percent, and in the retail trade sector from under 44 percent to over 65 percent. Similarly, the share of the top five groups by assets rose sharply by 2021 to 68 percent in the manufacturing of basic metals, 26 percent in the manufacturing of chemicals, 90 percent in the manufacturing of refined petroleum and coke, and 47 percent in the manufacturing of nonmetallic mineral products (including cement and other building materials).

Figure 1. Falling Industrial Concentration since 1991, but Recent Rise (Assets)



Sources: Centre for Monitoring Indian Economy (CMIE) Prowess Dx database and author's joint work with Rahul Singh Chauhan of the University of Chicago.

nonfinancial sectors in 2021.[19] Data from the CMIE show the following patterns. First, until 2010, the Big-5 increased their footprint in more and more industrial sectors, broadening their reach to forty NIC two-digit nonfinancial sectors (figure 2, panel A). After this breadth-first strategy came the depth-next strategy. Starting in 2015, the Big-5 started acquiring larger and larger share within the sectors where they were present (see figure 2, panel B). In particular, their share in total assets of the nonfinancial sectors rose from close to 10 percent in 1991 to nearly 18 percent in 2021, whereas the share of the next biggest five business groups fell from 18 percent in 1992 to less than 9 percent. In other words, the Big-5 grew not just at the expense of the smallest firms but also of the next-largest firms.[20] It is possible that some of this growth in share of the Big-5 is due to their ability to acquire relatively large defaulted companies that filed to the IBC following the RBI-initiated cleanup of the banking sector in 2017–2018. However, figure 2, panel B, reveals the growth in share of the Big-5 starts earlier, in 2015 and not in 2018 when the first IBC cases started being resolved.

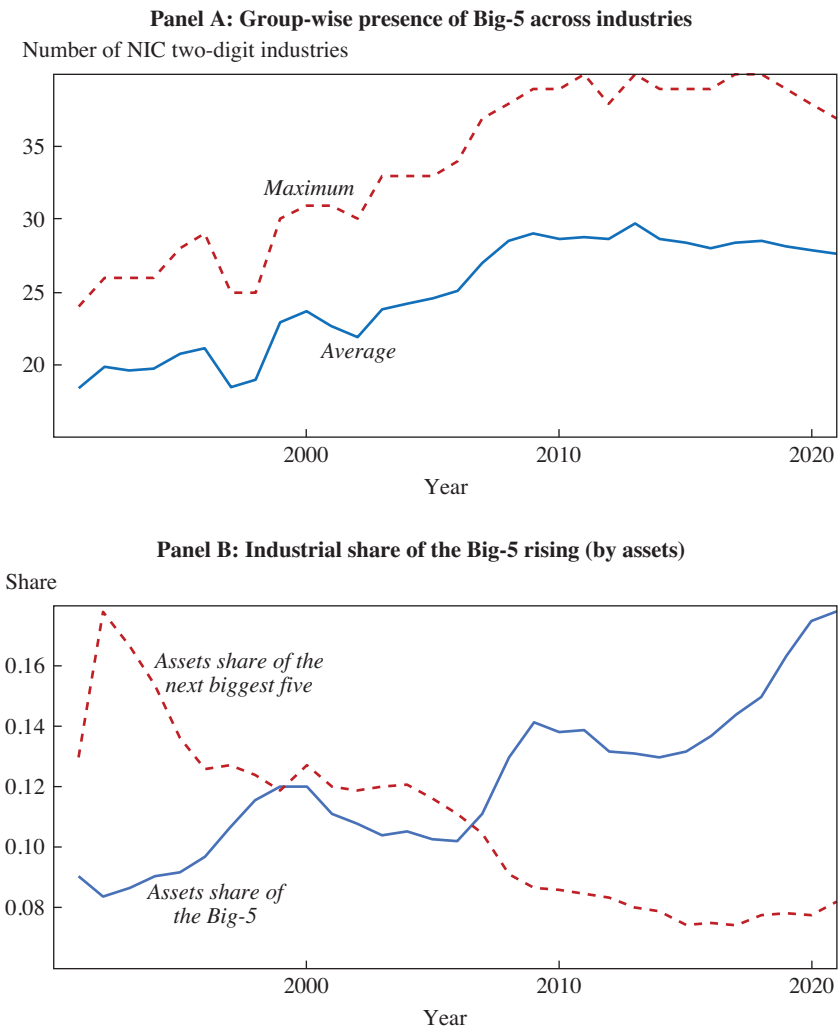
Next, this growth of the Big-5 appears to be driven in part by their growing share of overall mergers and acquisitions activity. Even though the aggregate number of merger and acquisition deals has dropped since 2011, the share of deals by the Big-5 has doubled from under 3 percent in 2015 to 6 percent in 2021, without such an increase being seen in the next five biggest groups (see figure 10 in the online appendix). Arguably, this growth has also been supported by a conscious industrial policy of creating "national champions" via preferential allocation of projects and, in some cases, regulatory agencies turning a blind eye to predatory pricing. Equally importantly, given the high tariffs, the Big-5 groups derive most of their revenues domestically as they do not have to compete with international peers in many sectors where they are present.

Such growth of conglomerates raises several concerns, such as the risk of crony capitalism, that is, political connections and inefficient project allocations, related party transactions within their byzantine corporate organization charts, over-leveraging due to an implicit perception that they are "too big to fail," key men or women (or key family) risk in their operational efficiency, and a lack of creative destruction by crowding out

19. The Big-5 are Reliance Group (Mukesh Ambani), Tata Group, Aditya Birla Group, Adani Group, and Bharti Telecom Group.

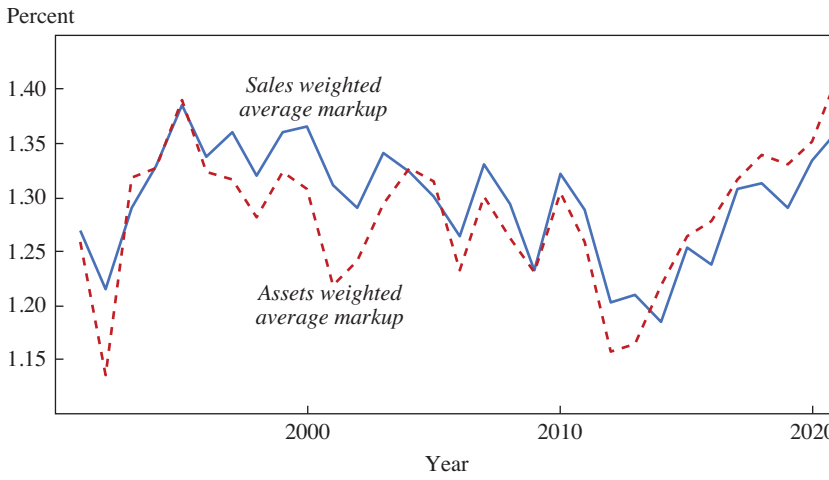
20. The largest contributing sectors to Big-5 sales are manufacturing of metals, manufacturing of coke and refined petroleum products, retail trade, and telecommunications. Prowess Dx data also confirm that, consistent with the rising market concentration, the Big-5 are receiving a greater percentage of their sales revenue from these sectors.

Figure 2. Rising Concentration of Big-5 by Presence across Industries and by Assets



Sources: Centre for Monitoring Indian Economy (CMIE) Prowess Dx database and author's joint work with Rahul Singh Chauhan of the University of Chicago.

Figure 3. Reversal in Markups



Sources: Centre for Monitoring Indian Economy (CMIE) Prowess Dx database and author's joint work with Rahul Singh Chauhan of the University of Chicago, using replication code from De Loecker, Eeckhout, and Unger (2020).

of entrants. The importance of these issues notwithstanding, I limit my attention below to how the rising market power of conglomerates affects product markups.[21]

Figure 3 shows that the rising industrial market power is indeed coincident with rising firm-wise markups since 2016. Specifically, we are interested in a measure of the markup which answers the question: If input cost of a firm changes by 1 percent, by how much does the product price change? Using the replication code provided by De Loecker, Eeckhout, and Unger (2020) for estimating firm-wise markups based on data from firm balance sheets, Rahul Singh Chauhan and I estimate firm-wise markups for Indian firms in the Prowess Dx database. Figure 3 illustrates the rise in aggregate sales-weighted and assets-weighted markups, which are barometers of market power in the nonfinancial industries. Markups fell gradually from the early 1990s until 2013 but have been rising significantly thereafter, even when capacity utilization in Indian industry was low during the pandemic

21. See, for example, the discussion in Philippon (2019) comparing and evaluating concentration trends between the United States and Europe and highlighting that "the great reversal" in the United States in industrial competition is raising firm markups and prices. This is an underappreciated—perhaps even easily dismissed—phenomenon by many economists.

due to collapse of aggregate demand and scaling in 2021 the high level of 1.4 seen in the 1990s.

While a deeper and fuller inquiry is warranted, we find that there is a potentially causal link from market power to markups. To illustrate the econometric results visually, figure 4, panel A, shows the industry-adjusted markups of the Big-5 and the rest of the firms, establishing a persistent and substantial 0.1–0.3 (i.e., 10–30 percentage points) markup gap between the two groups over the past two decades. Interestingly, there is no such robust pattern in figure 4, panel B, for the top five firms in each industry in a given year (as explained earlier, the top five in a given year overlap but do not fully coincide with the Big-5 as of 2021). In other words, it is the Big-5 that are able to exert extraordinary pricing power and capture economic rents relative to other firms in the industry, whereas the top five firms within an industry that are not in the Big-5 are not associated with such an outcome in markups.[22] By way of a concrete example, figure 18 in the online appendix shows the industry-adjusted markups of the Big-5 and the rest of the firms in manufacturing of nonmetallic minerals and basic metals. Since the 2000s, the Big-5 markup has been higher than the rest for nonmetallic minerals by 10–60 percentage points (a leading example being cement) and for basic metals by 10–20 percentage points. A similar wedge is observed in calculations for chemicals and chemical products.

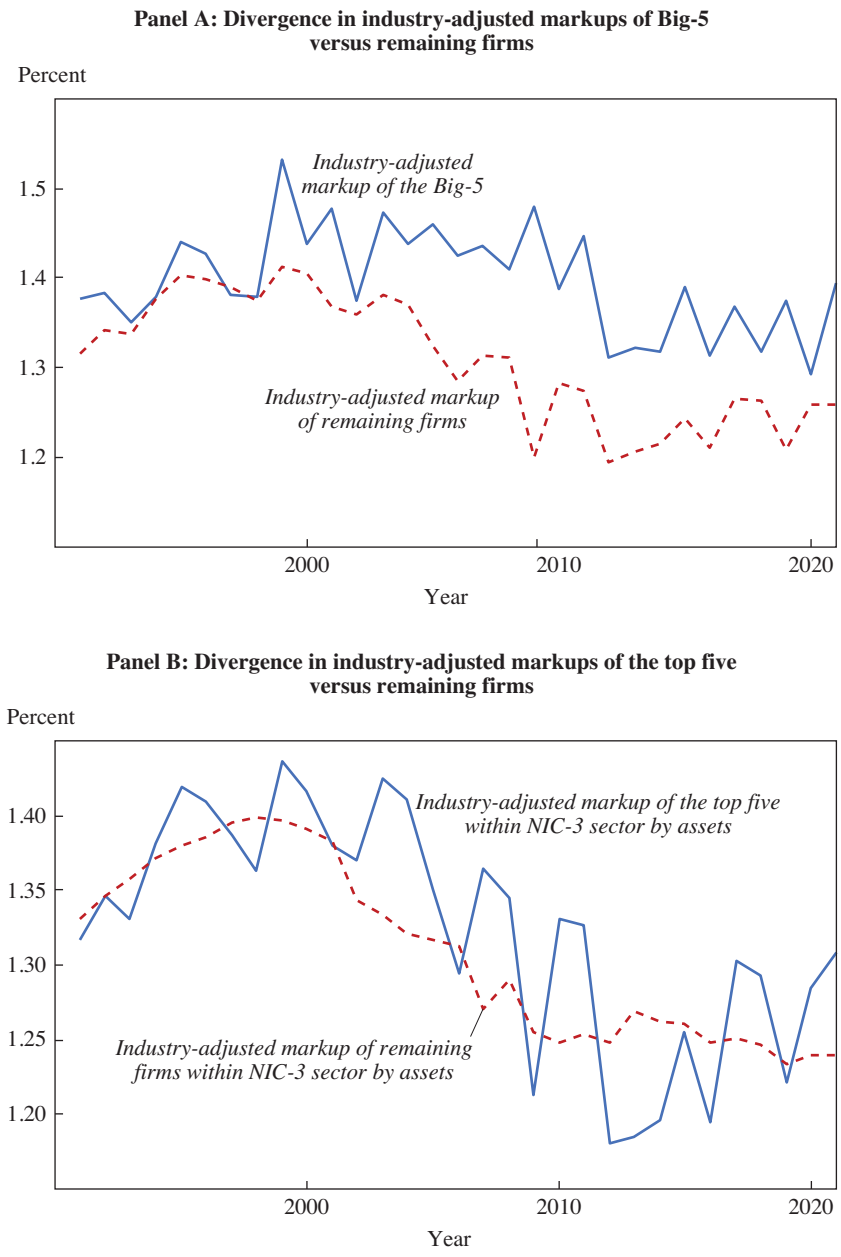
In summary, creating national champions, which is considered by many as the industrial policy of "new India," appears to be feeding directly into keeping prices at a high level, with the possibility that it is in turn feeding the persistent high level of India's core inflation (see III.D).

A natural question that arises is whether India is simply emulating the national champion policy of countries such as Korea, where large conglomerate groups (*chaebols*) such as Hyundai and Samsung have become significant international players in several sectors. There are at least two critical differences.

First, these countries did not protect their conglomerates with sky-high tariffs as India does. That is, their conglomerates were competing on a much greater level playing field with international peers, especially in the tradable sector, which likely explains their global competitiveness and the global brand status of several of their products. In contrast, barring

22. This differential pattern may be due to market power being driven by (1) the overall size of the group rather than the size of the industry-specific subsidiary, (2) larger size leading to better access to finance, (3) horizontally integrated position of the group in input-output matrices or supply chains, and (4) political patronage which may give credible comfort over future market share even while not lowering prices to compete for current share.

Figure 4. Divergence in Markups between Big-5, Top Five, and Other Firms



Sources: Centre for Monitoring Indian Economy (CMIE) Prowess Dx database and author's joint work with Rahul Singh Chauhan of the University of Chicago.

the exception of tech service exports, most of the revenues of India's Big-5 are domestically sourced, and barring the exception of e-commerce, without much foreign competition.

Second, these countries undertook a series of supply-side or factor-market reforms in land, labor, power, and financial sectors, among others.[23] These reforms made domestic competition vibrant. While India's financial sector has been restored to reasonable stability, critical reforms in land, labor, and power are either wanting or far from maturity. At present, the rising industrial concentration in India is safeguarded from both external competition via tariffs and domestic competition via poor access to factors of production. I contend, therefore, that the rising industrial concentration in India presents more of a risk or a dark side through the various distortions and inefficiencies flagged above rather than an opportunity or the bright side that could lead to the creation of globally competitive international giants.

III.C. Twin Deficits

India's stock of foreign exchange reserves is presently (March 2023) over $550 billion, which is more than twice the level at the time of the "taper tantrum" in 2013 and represents between eight and ten months of cover relative to its imports.[24] This has enabled the RBI to manage pretty well the exchange rate volatility in rupees since 2022 when the Federal Reserve embarked on a tightening of its monetary policy. Nevertheless, India's twin deficit metrics remain in what might be considered a less than comfortable zone.

The fiscal deficit, measured appropriately as a public sector borrowing requirement, that is, consolidating the center, the states, and their public sector enterprises, remains above 9 percent of GDP, and it has remained so for the past several years, reaching a peak of over 14 percent of GDP during 2020–2021 (Chinoy, Jain, and Sood 2022). In terms of outstanding stock, sovereign debt-to-GDP increased by 20 percent post-2020 and is presently close to 85 percent. Flow measures suggest an even greater concern as annual interest payments for the center are now over 30 percent of revenues and over 20 percent of expenditures (Mishra and Patel 2022), while real interest rates are rising in India and the rest of the world.

This has occurred in part because the targets set for fiscal deficit by the Fiscal Responsibility and Budget Management Act (FRBMA) of 2003

23. See Park (2002) on the East Asian model of economic development.

24. Reserve Bank of India, "Weekly Statistical Supplement," https://www.rbi.org.in/scripts/WSSViewDetail.aspx?PARAM1=2&TYPE=Section.

(reviewed by the NK Singh Committee set up in 2016) have been steadily missed under one pretext or another. The central government has done this in part to deliver on welfare, besides dealing with a weak economy, including during COVID-19. The state governments have contributed as well. Evaluated in a holistic sense, the states face severe hidden losses from the power sector, amounting to close to $40 billion annually, or 1.5 percent of GDP (Anand, Sharma, and Subramanian 2022) as power gets distributed at politically attractive prices rather than market prices, including and especially as a subsidy to agriculture. The resulting losses are perpetuated through state government balance sheets and national special purpose vehicles for the financing of the power sector—Power Finance Corporation (PFC) and REC (formerly Rural Electrification Corporation Limited). The debts of PFC and REC are generally not consolidated federally, requiring a focus on India's public sector borrowing requirement rather than just on balance sheet deficits. These statistics raise two significant risks.

One risk is that the fiscal dominance continues to hang like the sword of Damocles over the inflation-targeting and liquidity frameworks of the RBI, especially in politically important years. In turn, that makes inflation expectations hard to budge from post-COVID-19 highs. Historically, inflation has played a principal role in liquidating India's debts (Das and Ghate 2022), but it can take several years, even a decade, to do so. An equally unattractive or perhaps even worse alternative is that of financial repression, in which government-owned banks and insurance companies roll over national and subnational debts under moral suasion or under the guise of aggressive prudential norms. This crowds out private sector growth. One hopes that such repression will be harder to implement in a more market-dominated economy that India has now evolved into since 1990s. Fiscal dominance and financial repression nevertheless remain a threat.

Second, with such high fiscal deficits, there is a risk of crowding out long-term public expenditures in education and health. The Indian government deserves much credit for rationalizing subsidy (revenue) expenditures year after year toward public infrastructure (capital) expenditures and delivering welfare—including basic health services—more efficiently on the back of India's digital plumbing. However, this efficiency needs to be weighed against a crowding out of states from the tax base by the center, which has made the center's welfarism drive effective but reduced states' ability to incur capital expenditures, and which some view as a potent threat to the country's cooperative federalist structure.

Now, let us turn to India's current account deficit (CAD). While a sharp fall in commodity prices and a surge in tech and non-tech services exports

has brought the expected value of CAD to about 2 percent of GDP for 2023–2024, it averaged close to 3 percent for the period of March 2021 to September 2022 with several reported prints of the statistic in excess of 3 percent (Chinoy, Jain, and Sood 2022). This reflects India's poor share of goods exports in spite of excellence in services exports, its consumption being lopsided toward the urban households who consume several imported goods, and the sticky core inflation inducing greater gold imports. A corollary is that without broad-based consumption growth, India seems unable to grow at or close to its potential level. Every time it seeks to do so, it experiences merchandise trade deficits that raise the current account deficit, inducing exchange rate weakness and imported inflation (especially when oil prices rise). These spillovers, in turn, necessitate that the RBI has to raise interest rates to rein in inflation, creating an unavoidable dampener on any nascent investment cycle.

III.D. Persistent (Core) Inflation

Core inflation, that is, headline consumer price inflation excluding food and fuel components, has been persistent in India at around 6 percent during 2020–2022. Headline inflation has by and large hovered around the core—even gravitated toward it (Chinoy, Jain, and Sood 2022)—and steadfastly moved away from the inflation targeting mandate of 4 percent (±2 percent) while paying attention to growth. Alternative definitions of "core"—trimmed means and diffusion indexes—all suggest broad-based inflation is underway, reflecting in part the demand effect of post-COVID-19 stimulus, particularly in urban segments.

This may, however, not be the entire picture. There seems to be an urban wage spiral in the fastest growing sectors such as IT services where export demand remains high. Formally available statistics on listed company wage growth also appear in double digits, that is, definitively in excess of the inflation rate. Concomitantly and consequently, household and business inflation expectations have risen. As some analysts have noted (Chakraborty and Baqar 2023), the rise of core inflation and its persistence, as well as the urban wage spiral, are puzzling given the increasing slack in the overall employment scenario. Effectively, India's Phillips curve seems to have moved up or steepened as it seeks to close the post-pandemic output gap, reflecting lack of adequate skilled labor for the formal sector, which is outperforming the informal sector but which is unable to penetrate or upgrade the labor force.

Another reason why core inflation is persistently rising is that consumers do not seem to be fully benefiting from input price declines, which

may be due to greater pricing power in increasingly concentrated industrial organization structures (see III.B). What lends some credibility to this thesis is the observation that, in contrast to the rest of the world, India's core inflation is rising more in goods, where its industrial sectors are increasingly concentrated, than in services, though there are early signs of pricing power rising in the services sector too.

III.E. Education Gaps, Declining Female Labor Participation, and Too Few Jobs

Finally, the substantial subsidization of input factors (electricity, fertilizers, water, credit) to the agricultural sector keeps the sector artificially large. As per data from the World Bank, while India's agricultural labor force share has shrunk from 64 percent in 1991 to 44 percent in 2020, it remains way too large in an absolute sense given that the share of agriculture in GDP during 2015–2020 has been in the 16–18 percent range (compared to just 4 percent for the rest of the world).[25] Further, the sector operates at low efficiency in that the value-add per Indian agricultural worker is only 8 percent on a unit of investment.[26] Overall, this has kept the distribution of workers low-skilled and unfavorable for India being able to grow services exports to their full potential without immediately inducing a wage spiral.

The chicken-and-egg problem is hard to resolve, but labor persisting in low-skilled jobs is consistent with substantial education gaps for the development of high-skilled labor. This is in spite of a steady improvement in school enrollments in India since 2006 (as per the *Annual Status of Education Reports*), including for girls. In particular, literacy levels have dropped steadily over the past decade: reading ability is presently below the pre-2012 levels, in both government and private schools and for both boys and girls; arithmetic levels have dropped less steeply but are presently at lower levels than in 2018 (ASER Centre 2023).

While some of the education gaps are undoubtedly due to extended school closures during the COVID-19 lockdowns and beyond, perhaps the biggest impact of the pandemic years has been on India's female labor

25. The World Bank, "Agriculture, Forestry, and Fishing, Value Added (% of GDP)—India," https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?locations=IN; "Employment in Agriculture (% of Total Employment) (Modeled ILO Estimate)," https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS.

26. Statista, "Gross Value Added (GVA) per Worker in Agriculture in India from Financial Year 2016 to 2023," https://www.statista.com/statistics/1131117/india-gross-value-added-in-agriculture-per-worker/.

force participation. As per survey data from the CMIE, it has declined from 18 percent in 2016 to under 11 percent in 2022, and somewhat unexpectedly, to under 7 percent in urban areas.[27] These trends hint strongly at the lack of adequate job creation in the aggregate, seen for instance in the CMIE survey data.

Finally, formal job creation measured using Employees' Provident Fund Organisation data (pension fund enrollment data) showed a decline of 15.5 percent in November–December 2022 in the 18–25 age group of typically new subscribers (Kaul 2021; Rajora 2023). Consistently, the survey statistics of the CMIE show that the unemployment rate has risen from 3.37 percent in July 2017 to 6.9 percent in January 2021 (it being greater than 20 percent for graduates and postgraduates) to 7.5 percent in March 2023. In response, the labor participation rate has also declined from 49 percent in July 2016 to 41 percent in January 2021. This scarcity of jobs in the patriarchal Indian society has left women effectively out of the labor force. Unsurprisingly, there has also been a substantial reduction in salaries of women who are in the labor force and employed.

Overall, India seems to be creating too few jobs relative to its labor force needs, there are too many low-skilled laborers (especially in agriculture), primary education gaps are mounting, and the female labor force is bearing the brunt of many of these developments.

IV. Proposals

The challenges India is presently saddled with provide a natural blueprint of some structural reforms that can be initiated immediately for reaping rewards over the next decade.

IV.A. De-tariff

India's import tariffs are way too high and protectionist in favor of its incumbents. India needs to bring tariffs in line at least with those of China and perhaps, to have an advantage over its key competitors, make them even lower. If a onetime sharp drop in high tariff rates is difficult, for example, for managing the re-skilling of displaced labor, then policy can announce a calibrated reduction plan over, say, a three-year period. Such clarity of purpose and forward guidance would in itself facilitate the

27. Even data from the Periodic Labor Force Survey (PLFS) show that female labor force participation was at 27 percent in 2021–2022, which is low in an absolute sense as well as relative to most peer countries.

expansion of goods trade, induce a much-needed global competitiveness in its manufacturing firms, and likely also encourage an uptick in investment.

I stress here that tariffs in agriculture need to be reduced too, given their much higher levels than in other sectors. The reductions will have to be sharper to start with but persisted with in order to enable this sector to downsize in its labor share and upgrade in its efficiency. Further, India being more in line with international tariff rates might facilitate its greater participation in trade agreements, serving in turn as a pre-commitment not to raise tariffs again.[28]

IV.B. Dismantle or Reduce the Market Power of Indian Conglomerates

How should India move away from the rising industrial concentration? Several risks have materialized in a rather short period of time during 2023 in the case of one of the largest (Big-5) conglomerates, and de-leveraging may slow down medium-term investments by this conglomerate. Given this experience, it is worthwhile to entertain mechanisms for reducing the market power of the largest conglomerates.

As seen in figure 2, panel A, the Big-5 conglomerates are in more than forty NIC two-digit sectors. Hence, one way out of their breadth of presence is the good old Theodore Roosevelt or William Howard Taft style trustbuster strategy of simply breaking up large industrial firms and their monopolies or oligopolies by regulatory fiat or via competition commission diktat. This has been done repeatedly in the United States when the concentration of corporate power has risen nationally in a sector or across different product lines. One advantage of this approach is that it would require various resulting subgroups to have separate—and likely more transparent—balance sheets as well as ownership, management, and governance structures. Such trustbusting may, however, be awkward for the current government given it has—by revealed preference—adopted an industrial policy favoring national champions.

An alternative route would be to throw sand in the wheels of rising industrial concentration by making it economically unattractive to remain a large conglomerate unless productivity gains are truly large. As discussed earlier, the Big-5 have grown their market share over the past decade by increasing their footprint in mergers and acquisitions. It could be required

28. India's public food stockpiling program, for example, is intended as a food-price security system, acts to support agricultural income in normal times but depresses it when prices rise, leads to routine restrictions on imports and exports of specific agricultural items, and causes India to disrupt WTO negotiations even in areas ostensibly unrelated to agriculture and food security.

that they own more equity of the companies they acquire, for example, 80 percent or higher as in the United States (Morck 2005; Kandel and others 2019), in order to get benefits that group companies enjoy. These benefits typically include tax exemption on dividends from subsidiaries to parents, consolidation of income between subsidiaries and parents for tax purposes (generally beneficial due to offsetting of losses against profits), and tax exemption on the spin-off of subsidiary shares to parent shareholders. In essence, by increasing the extent of subsidiary ownership required for earning the benefits of being a conglomerate, some of the conglomerates may spin off existing subsidiaries where gains do not justify such an increase.

Whether done by brute force as a competition commission diktat or more gracefully, it would be to India's advantage to become more competition-friendly and less incumbent-friendly, especially less conglomerate-friendly. A significant benefit would be that even if the subgroups remain among the largest companies (say, the top five) in their respective sectors, they may lack the pricing power commanded by the Big-5 (an important difference illustrated in figure 4).

IV.C. Get Insolvency and Bankruptcy Code (IBC) Back on Track

While the deterrence effect of IBC is well at work in the form of lower incidence of defaults, the progress in resolving cases through bankruptcy is slow, which adds to the substantial erosion of asset and franchise values of defaulted companies. The legal benches handling the cases and the intermediate steps leading to the eventual reorganization, sale, or liquidation of the defaulted companies could be subject to a tighter timeline, closer to that originally envisaged under the IBC. The present average of resolution times, which is nearly eighteen to twenty-four months, seems appropriate only for the largest of the cases and in difficult economic times. Most other cases should be resolved much faster. One possibility is that many small and frivolous bankruptcies can be resolved privately outside of the IBC to prevent choking the pipeline of cases. The Indian Banks' Association could consider templates for bank-loan trip-wire covenants that can trigger such early resolution via prepackaged bankruptcies.

Separately, a true stress test for the IBC would be whether it can handle well a large conglomerate's default, either at the group level or at one or more of its subsidiaries. Going by the market credit spreads on the internationally issued bonds of some of these conglomerates, this is not at all outside the realm of reasonable probability. Resolution of such entities is not entirely unlike resolving a large, complex financial institution or a

systemically important one. Should India's large conglomerates be subject to a living will or resolution planning requirement, as required of the systemically important financial institutions in many parts of the world? Design of such living wills may also lay bare their complex web of related party transactions and create an indirect tax on being large for rent-seeking rather than productivity gains.

IV.D. Deliver on the FRBMA (Fiscal Deficit) Targets

Over the past decade, the fiscal deficit targets were first missed slowly and then simply kept in abeyance. A credible glide path needs to be provided to bring realized deficits in line with these targets. Clearly, higher growth from rationalization of revenue (subsidy) expenditures toward capital expenditures is one way to achieve this, but as this transition tends to be slow, will it be sufficient? The central government can use the presently buoyant tax collection phase to glide faster to targets. What matters in the end though is the overall public sector borrowing requirement. On this front, power sector and distribution companies' woes seem important to address. As mentioned earlier, their losses seem to be on the order of 1.5 percent of GDP on an annual basis. A first step would be to create a national grid for the power market to allow efficient use of capacity and market pricing based on that. A second step could be to create a time-bound transition to rationalizing the highest of the subsidies and leakages (such as for rural electrification and in agriculture) that result in substantive losses. Finally, some states will gain and others lose in the process, and the central government could create a burden-sharing mechanism to redistribute gains and losses across states. Implementing such a step successfully requires visionary leadership and can help restore confidence in India's cooperative federalism compact between center and states.

IV.E. Deliver on the Inflation Mandate

The RBI—even if reluctantly, and in all likelihood induced by the Federal Reserve's tightening actions—has shown commitment since May 2022 to bringing headline inflation in line with the mandated target of 4 percent and not simply at or above the upper band limit of 6 percent. However, the RBI's task has been rendered difficult by the persistence of elevated core inflation, which as of March 2023 hovered around 6 percent (Chinoy, Jain, and Sood 2022). It is a reasonable assumption that headline inflation will eventually veer toward the core. Surprises in food prices, due to uncertain monsoons (e.g., the risk from El Niño), and in the price of oil, if the Russian invasion of Ukraine remains unresolved, are likely to hike

inflation up further. Hence, in my view, the best the RBI can do is to invest extra in inflation-targeting credibility by raising real rates further and sacrificing some growth in the short term (I stress *only* in the short term). Such sacrifice seems crucial to bring investor and business inflation expectations down and arrest the upward wage spiral in the formal sector on a durable basis. Gains from inflation targeting in terms of lowering inflation expectations durably might never get fully realized if the central bank is seen as routinely sacrificing its inflation target for supporting growth, as was necessary at the time of COVID-19, but is not seen as determined in the other direction.

IV.F. Address Skill and Education Gaps

There are three critical steps I suggest India undertake on this front.

First, the share of low-earning agricultural labor needs to be reduced over time and be transformed into better-skilled, higher-earning manufacturing and services labor. While that requires creating more jobs in the non-agricultural sectors, it also requires a willingness on the part of young labor to leave the agricultural sector. One way is to raise the sector's presently subsidized costs of inputs to market levels over a period of time, allow foreign entry into the sector and lift its productivity by lowering tariffs, and have a plan to retrain a part of the labor force—effectively lowering entry rate into the sector by training the youth—for vocational skills in manufacturing and services. This could be taken on as a flagship project by the ministry dedicated to skills development in partnership with private firms.

Second, the huge primary education gaps created in children's learning all over India during the pandemic need to be addressed in a decisive manner. While there are many initiatives that could do the needful, one option is to deliver a grade-by-grade national curriculum for a thirty-day remedial summer program and another enriched thirty-day start-of-the-year boot camp for reinforcement. Municipal schools can be required to adopt the programs mandatorily. Private schools may join voluntarily if the curriculum is attractive. ASER-style surveys could be conducted pre-summer, end-of-summer, and on completion of boot camp to assess success, identify where gaps remain, and plan next steps accordingly for further remedial action.

Related to general education, it has always struck me that India does not have charter or magnet public schools providing the highest-quality science, technology, engineering, and mathematics (STEM) education at middle school and high school levels. Such schools could be set up in each state, with screening based on entrance tests, in order to create an

aspirational learning path among most of India's less privileged children who go to average or below-average municipal schools. Essentially, this can be modeled along the lines of the Indian Institutes of Technology but for primary and secondary education. Long-term payoffs would be substantial. The experience of similar programs in New York and Massachusetts in the United States offers a possible role model for execution (Angrist, Pathak, and Walters 2013).

Finally, it is important to make it easier for women to join the labor force, especially in urban areas where the fall in their participation rate has been the highest. Given that Indian companies are required to contribute a minimum of 2 percent of their net profits over the previous three years for corporate social responsibility, the following could be made qualifying for such expense: (1) support for entities—not-for-profit or otherwise—that invest in education of girls and skill development of the young female population, including companies' own initiatives; (2) suitable leaves for maternity and paternity that can be substituted between spouses, so as to increase the flexibility women have in resuming work after bearing children; and (3) setting up of quality childcare facilities on company premises or in neighborhoods to reduce the domestic burdens of working women. Similar schemes can be worked out for rural areas with partner organizations and with some public, multilateral, or large NGO financial support.[29]

While this is a tall order, it may be the imperative to fully realize India's demographic dividend.

ACKNOWLEDGMENTS I am grateful to Janice Eberly, Maury Obstfeld, and James Stock for inviting this piece; for inputs from Swaminathan Aiyar, Yakov Amihud, Tamal Bandyopadhyay, Pranjul Bhandari, Olivier Blanchard, Sajjid Chinoy, Kaushik Das, Kristin Forbes, Ashish Gupta, Nikhil Gupta, Dharmakirti Joshi, Vivek Kaul, Amartya Lahiri, Neelkanth Mishra, Prachi Mishra, N. R. Prabhala, Rupa Rege Nitsure, Raghuram Rajan, Ken Rogoff, Kush Shah, Duvvuri Subbarao, Arvind Subramanian, Y. V. Reddy, and seminar participants at the Spring 2023 *BPEA* Conference and the March 2023 NYU-NSE Initiative on Indian Capital Markets conference; and for excellent research assistance by Ridhika Agrawal and Rahul Singh Chauhan.

29. This may also support attitudinal shifts away from India's patriarchal past, which are crucial to ensure that trade liberalization promotes gender equality rather than hurting it, as was the case with the 1991 reforms (Gupta 2021).

References

Acharya, Viral V. 2020. *Quest for Restoring Financial Stability in India*. New Delhi: SAGE India, 2020; second edition, New Delhi: Penguin India, 2023.

Aiyar, Swaminathan S. Anklesaria. 2018. "India's New Protectionism Threatens Gains from Economic Reform." Policy Analysis 851. Washington: Cato Institute.

Anand, Abhishek, Navneeraj Sharma, and Arvind Subramanian. 2022. "Power, a Reality Check." *Indian Express*, August 18. https://indianexpress.com/article/opinion/columns/power-a-reality-check-analysis-of-state-government-finances-8094214/.

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. 2013. "Explaining Charter School Effectiveness." *American Economic Journal: Applied Economics* 5, no. 4: 1–27.

ASER Centre. 2023. *Annual Status of Education Report 2022*. New Delhi: Author. https://asercentre.org/aser-2022/.

Bhandari, Pranjul, and Aayushi Chaudhary. 2022. *India: From Scars to the Stars?* London: HSBC Global Research.

Chakraborty, Samiran, and Baqar, Z. 2023. *India Economics: Why Is Core Inflation So Persistent?* Citi Research.

Chan, Szu Ping. 2019. "Why India Is One of World's Most Protectionist Countries." BBC News, April 11. https://www.bbc.com/news/business-47857583.

Chatterjee, Shoumitro, and Arvind Subramanian. 2020. "India's Inward (Re)Turn: Is It Warranted? Will It Work?" Working Paper. Social Science Research Network, December 1. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3713253.

Chinoy, Sajjid Z., Toshi Jain, and Divyanit Sood. 2022. *India in 2023: Balancing Growth with Stability*. Mumbai: JPMorgan Chase Bank.

Das, Piyali, and Chetan Ghate. 2022. "Debt Decomposition and the Role of Inflation: A Security Level Analysis for India." *Economic Modelling* 113:105855.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications: Econometric Tools for Analyzing Market Outcomes." *Quarterly Journal of Economics* 135, no. 2: 561–644.

Dubey, Tamanna Singh, and Amiyatosh Purnanandam. 2023. "Can Cashless Payments Spur Economic Growth?" Working Paper. Social Science Research Network, March 4. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4373602.

Gupta, Ashish, and Viral Shah. 2021. *Indian FinTech Sector: A Guide to the Galaxy*. Zurich: Credit Suisse Research.

Gupta, Ashmita. 2021. "Effect of Trade Liberalization on Gender Inequality: The Case of India." *IMF Economic Review* 69, no. 4: 682–720.

Gupta, Manish, Nitesh Jain, Gouri Prasad Panda, Gaurav Chaturvedi, Aditi Phatak, Rajesh Kumar Singh, Kushagra Joshi, Vivek Tiwari, and Mayank Kanojia. 2021. *Reinforcing the Code*. Mumbai: CRISIL Research and New Delhi: Associated Chambers of Commerce and Industry of India. https://www.crisil.com/content/dam/crisil/mailers/rating-newsletter/2021/september/reinforcing-the-code.pdf.

Kandel, Eugene, Konstantin Kosenko, Randall Morck, and Yishay Yafeh. 2019. "The Great Pyramids of America: A Revised History of U.S. Business Groups, Corporate Ownership, and Regulation, 1926–1950." *Strategic Management Journal* 40, no. 5: 781–808.

Kaul, Vivek. 2021. "Female Labour Participation Rate: Revealing the Real Picture behind India's Unemployment Problem." March 8. https://vivekkaul.com/tag/female-labour-participation-rate/.

Kulkarni, Nirupama, S. K. Ritadhi, Siddharth Vij, and Katherine Waldock. 2019. "Unearthing Zombies." Working Paper. Social Science Research Network, December 16. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3495660.

Mishra, Neelkanth, Abhay Khaitan, Varun Ahuja, Krati Sanklecha, Anubhav Aggarwal, Lokesh Garg, and others. 2021a. *100 Unicorns: India's Changing Corporate Landscape*. Zurich: Credit Suisse Research.

Mishra, Neelkanth, Abhay Khaitan, Varun Ahuja, Krati Sanklecha, Anubhav Aggarwal, Lokesh Garg, and others. 2021b. *Unicorns: After a Sprint, a Pause for Breath*. Zurich: Credit Suisse Research.

Mishra, Prachi, and Nikhil Patel. 2022. "The Anatomy of Public Debt Reductions in India." Working Paper. Philadelphia: University of Pennsylvania Center for the Advanced Study of India. https://casi.sas.upenn.edu/iit/prachimishra-nikhilpatel.

Mohan, Rakesh. 2021. "The Response of the Reserve Bank of India to COVID-19: Do Whatever It Takes." In *Monetary Policy and Central Banking in the COVID Era*, edited by Bill English, Kristin Forbes, and Ángel Ubide. London: Centre for Economic Policy Research.

Mohommad, Adil, Mehdi Raissi, Kyuho Lee, and Chanpheng Fizzarotti. 2023. "Volatile Commodity Prices Reduce Growth and Amplify Swings in Inflation." Blog post, March 28, International Monetary Fund. https://www.imf.org/en/Blogs/Articles/2023/03/28/volatile-commodity-prices-reduce-growth-and-amplify-swings-in-inflation.

Morck, Randall. 2005. "How to Eliminate Pyramidal Business Groups: The Double Taxation of Intercorporate Dividends and Other Incisive Uses of Tax Policy." *Tax Policy and the Economy* 19:135–79.

Park, Jong H. 2002. "The East Asian Model of Economic Development and Developing Countries." *Journal of Developing Societies* 18, no. 4: 330–53.

Philippon, Thomas. 2019. "The Economics and Politics of Market Concentration." *NBER Reporter*, no. 4. https://www.nber.org/reporter/2019number4/economics-and-politics-market-concentration.

Rajora, Shiva. 2023. "Fresh Formal Job Creation below 1 Million for Third Month: EPFO Data." *Business Standard*, February 20. https://www.business-standard.com/article/economy-policy/fresh-formal-job-creation-remains-below-1-mn-mark-for-third-straight-month-123022000966_1.html.

Reserve Bank of India (RBI). 2018. *Financial Stability Report: Issue No. 18*. Mumbai. https://rbidocs.rbi.org.in/rdocs//PublicationReport/Pdfs/0FSRDECEMBER2018DAFEDD89C01C432786925639A4864F96.PDF.

Reserve Bank of India (RBI). 2022. *Financial Stability Report: Issue No. 26*. Mumbai. https://rbidocs.rbi.org.in/rdocs//PublicationReport/Pdfs/0FSRDECEMBER2022F93A2F188A394ACDB2FDDC2FCC0D07F0.PDF.

Sinha, Shreyas. 2023. "Can India Become the Next Superpower?" Fifty50, February 17. https://fifty50politics.substack.com/p/fifty50-can-india-become-the-next.

Totty, Michael. 2022. "Addressing Its Lack of an ID System, India Registers 1.2 Billion in a Decade." UCLA Anderson Review, April 13. https://anderson-review.ucla.edu/addressing-its-lack-of-an-id-system-india-registers-1-2-billion-in-a-decade/.

Vishwanathan, N. S. 2018. "It Is Not Business as Usual for Lenders and Borrowers." Speech given at the National Institute of Bank Management, Pune, April 18. https://www.bis.org/review/r180420e.pdf.

CARLOS CARVALHO
Kapitalo Investimentos and PUC-Rio

FERNANDA NECHIO
Federal Reserve Bank of San Francisco

Challenges to Disinflation: The Brazilian Experience

ABSTRACT We review two previous bouts of high inflation and disinflation since Brazil adopted inflation targeting. In both episodes, fiscal sustainability concerns were present and inflation expectations became unanchored despite substantial monetary policy tightening. Disinflation and the reanchoring of expectations took time and proved costly, as both episodes entailed a recession. They required tight monetary policy combined with critical shifts toward structural economic reforms and sound fiscal policy. The ongoing episode features the same fiscal concerns and unanchored inflation expectations. This suggests the path ahead for disinflation will be challenging, unless policies change direction. We also speculate whether the Brazilian experience can provide insights for other countries.

In response to the COVID-19 pandemic, governments around the world pursued unprecedented public health, fiscal, and monetary policies. To varying degrees, such policies contributed to supply and demand imbalances that led to a sharp increase in inflation worldwide. While the rise of inflation was initially concentrated in a few sectors, such as food and energy, various measures of core inflation show it has become more widespread

Conflict of Interest Disclosure: Fernanda Nechio was deputy governor at the Central Bank of Brazil between July 2019 and May 2021. Carlos Carvalho was deputy governor of the Central Bank of Brazil between July 2016 and September 2019. 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. The authors are not currently an officer, director, or board member of any organization with a financial or political interest in this paper.

Brookings Papers on Economic Activity, Spring 2023: 217–241 © 2024 The Brookings Institution.

over time. The war in Ukraine was an additional inflationary shock, with particularly acute effects in Europe.

The unique nature of the pandemic and associated policy responses made it challenging to anticipate how economies would behave. Sectoral demand shifts, supply disruptions, and the dynamics of social distancing or lockdowns and subsequent reopening made it particularly hard to anticipate inflation developments. While a few commentators got it more right than wrong, it is fair to say most market participants and policymakers failed to foresee how high and persistent inflation would become.

Now that many central banks have done the lion's share of policy tightening, economies show some signs of slowing, and inflation may have peaked, attention has turned to the awaited disinflation process. How fast is it going to be? How costly will it be in terms of output? At the time of writing, financial markets in the United States seem to reflect a benign view, whereby inflation will exhibit low persistence and fall with little cost in terms of foregone output—that is, a low sacrifice ratio. How likely is that outcome? The academic literature suggests such a benign, "soft landing" scenario is not very likely, although by no means impossible.[1]

The fact that the United States and other advanced economies have not experienced such high levels of inflation in decades, however, adds to the difficulties in drawing conclusions from historical experience. Especially to the point, most available estimates of the costs of disinflation in advanced economies pertain to periods prior to the adoption of modern inflation targeting regimes currently in place in many countries.

In contrast, some emerging economies have experienced more frequent episodes with comparably high levels of inflation, including episodes since they adopted inflation targeting. For those countries at least, perhaps their experiences can offer some insight about current disinflation challenges. Certainly subject to many additional caveats, they can complement the (also

1. Friedman and Schwartz (1963) and Romer and Romer (1989) provide a historical account of disinflation episodes in the United States and output costs of contractionary monetary policies. In an effort to quantify the output costs of disinflation, Ball (1991) discusses the roles of price rigidity and credibility problems, while Ball (1994) finds the sacrifice ratio is decreasing in the speed of disinflation and in the degree of wage flexibility. Sargent (1982, 2013) discusses the role of policy credibility in determining the costs of disinflation. Tetlow (2022) relies on dynamic general equilibrium models to provide a range of estimates of the costs of disinflation. More recently, Cecchetti and others (2023) rely on US and international evidence to conclude that sizable central bank–induced disinflation episodes are, more frequently than not, associated with economic downturns. Finally, Blinder (2023) offers a more optimistic view on the Federal Reserve's ability to achieve a soft landing.

imperfect) lessons about disinflation in advanced economies one can draw from the academic literature.

In this paper, we explore the case of Brazil, which adopted inflation targeting in 1999, after the collapse of its managed exchange rate regime (Bacha 2003). The current bout of high inflation is the third episode since then in which inflation reached levels significantly above target, and the Banco Central do Brasil (BCB) had to tighten policy significantly to rein in inflation.

The two previous episodes have a few features in common. In both cases, heightened fiscal concerns were present, the Brazilian currency depreciated significantly, inflation increased way above target, inflation expectations became or were already unanchored, and the BCB had to tighten policy significantly.

The two episodes ended with a sizable, albeit costly, disinflation. Both reversals involved many ingredients. Crucially, the course of fiscal policy was a positive surprise, the currency reversed its depreciation, and inflation expectations reanchored. Both episodes, however, produced a recession.

The current challenges for disinflation in Brazil feature the same credibility concerns regarding fiscal and monetary policies observed in the previous episodes. Analysts and market participants fear lack of fiscal discipline and inflation expectations have become unanchored. These similarities strongly suggest a successful disinflation will require a change of direction in economic policies going forward.

The Brazilian experience may be informative for challenges faced by other emerging economies where fiscal sustainability is a concern and where inflation expectations suggest monetary policy credibility is at stake. When it comes to advanced economies, analogies are much more imperfect. These countries have stronger fundamentals, much larger fiscal space, and a longer history of low and stable inflation and macroeconomic stability. Nevertheless, the behavior of long-run inflation expectations in the United States in the aftermath of the global financial crisis and during the recent confidence scare in the United Kingdom suggests policy credibility may be challenged even in advanced economies.

The rest of the paper is organized as follows. Section I describes the adoption of the inflation targeting regime in Brazil and identifies three periods of high inflation and unanchoring of inflation expectations since then. Section II provides an account of the first two high-inflation episodes. In section III we describe the current bout of high inflation and highlight lessons from the two previous episodes for current disinflationary efforts. Section IV concludes.

I. The Inflation Targeting Regime in Brazil

After a decade battling extremely high or hyperinflation, the Real Plan was successful in bringing down inflation in 1994.[2] At first, between 1994 and 1999, the country introduced a managed exchange rate regime that helped stabilize inflation. Fiscal policy remained loose, however, which eventually led to a confidence crisis by the end of 1998.

After abandoning the currency peg in January 1999, Brazil transitioned to an inflation targeting regime. Under this framework, the BCB sets the main policy rate (the Selic rate) aiming to maintain annual national consumer price inflation (as measured by the IPCA, the Extended National Consumer Price Index) on target. A tolerance band allows the BCB to accommodate shocks. The target and tolerance band are set three years in advance by the National Monetary Council, which is currently composed of the minister of finance, the minister of planning, and the BCB governor.

After an initial period with declining targets that were revised twice in response to the first bout of high inflation, the target remained at 4.5 percent between 2005 and 2018. Since 2019, the target was reduced by a quarter of a percentage point each year, until it reached 3 percent, currently in place for 2024 and 2025. Figure 1 highlights three instances in which inflation rose significantly above the tolerance band—around 2002, 2015, and, more recently, since 2021.

Of course, a proper assessment of an inflation targeting regime and the costs associated with bringing inflation back to target must take into account the dynamics of inflation expectations. When expectations are well anchored, high inflation episodes are more easily reversed, as agents continue to make their economic decisions assuming inflation will fall back to "normal" levels (i.e., around target). In contrast, when expectations are unanchored, agents anticipate that inflationary shocks will propagate and change their behavior accordingly.

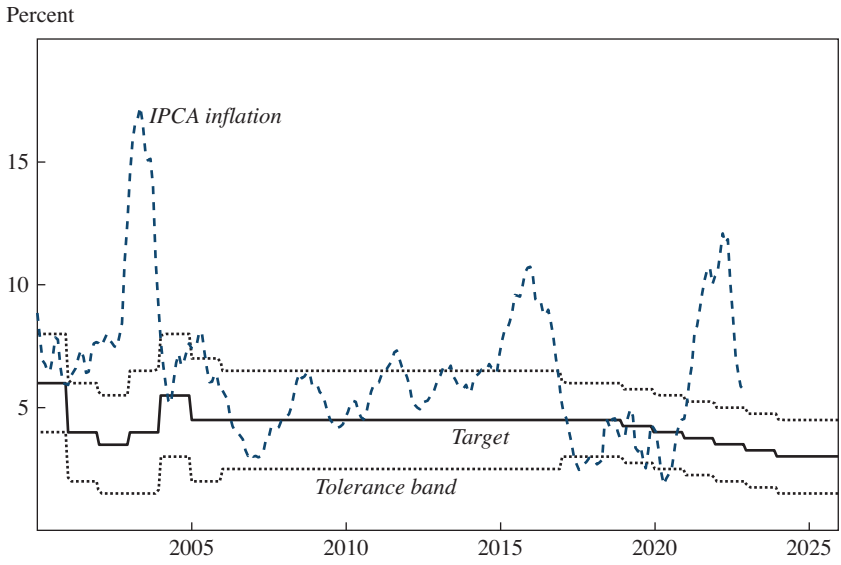
Therefore, before turning to an account of the first two high inflation episodes, we discuss the dynamics of expectations and their degree of anchoring since Brazil adopted inflation targeting.

I.A. Inflation Expectations and the Degree of Anchoring

To closely follow market participants' views on economic conditions, in 2002 the BCB started a survey of professional forecasters, known as the

2. For a detailed account of the Real Plan, see Bacha (2003).

Figure 1. Inflation Targeting Regime in Brazil



Source: Banco Central do Brasil.

Focus Survey.[3] Currently, it has around 140 participants, of which 100 provide forecasts frequently. The survey covers a range of macroeconomic variables, such as inflation, the Selic policy rate, the exchange rate, various fiscal variables, economic activity, and balance of payments. Participants can update their forecasts for various horizons daily if they so wish. In order to provide incentives for participants to report frequently and carefully, the BCB runs a few forecasting contests and publishes the five most accurate forecasters. The BCB also releases aggregate survey statistics, such as means, medians, and dispersion measures.

Abib and others (2023) use the Focus Survey to construct a measure of the degree of expectations unanchoring in Brazil, based on the credibility index introduced by Cecchetti and Krause (2002).[4] They measure the degree of expectations unanchoring by taking deviations from

3. Banco Central do Brasil, "Market Expectations—Focus Survey," https://www.bcb.gov.br/en/monetarypolicy/marketexpectations.

4. For a discussion of inflation and central bank credibility in emerging economies, see Ha, Kose, and Ohnsorge (2019) and references therein.

target of twelve-month inflation expectations between two and three years ahead and converting them into an index with values between zero and one. Since the BCB considers its relevant monetary policy horizon to be around eighteen months ahead, the chosen horizon is arguably long enough to reflect policy credibility considerations rather than the delayed effects of ongoing shocks.[5] The authors validate this measure of the degree of unanchoring by looking at other measures the literature has proposed, such as the cross-sectional dispersion of expectations and variation in the extent of pass-through of short-run to long-run inflation expectations. In addition, using Brazilian Producer Price Index (PPI) micro price data, they find exchange rate pass-through into prices increases substantially when expectations are unanchored. We update their index and use data from January 2, 2002, to January 18, 2023.

The degree of unanchoring of inflation expectations, U_t, is given by:

$$U_t = \begin{cases} 1 & \text{if } \pi^{e}_{t+2,3} > \pi^{\max}_{t+2,3} \\ \dfrac{\pi^{e}_{t+2,3} - \pi^{*}_{t+2,3}}{\pi^{\max}_{t+2,3} - \pi^{*}_{t+2,3}} & \text{if } \pi^{*}_{t+2,3} \leq \pi^{e}_{t+2,3} \leq \pi^{\max}_{t+2,3}, \\ 0 & \text{if } \pi^{e}_{t+2,3} < \pi^{*}_{t+2,3}, \end{cases} \tag{1}$$

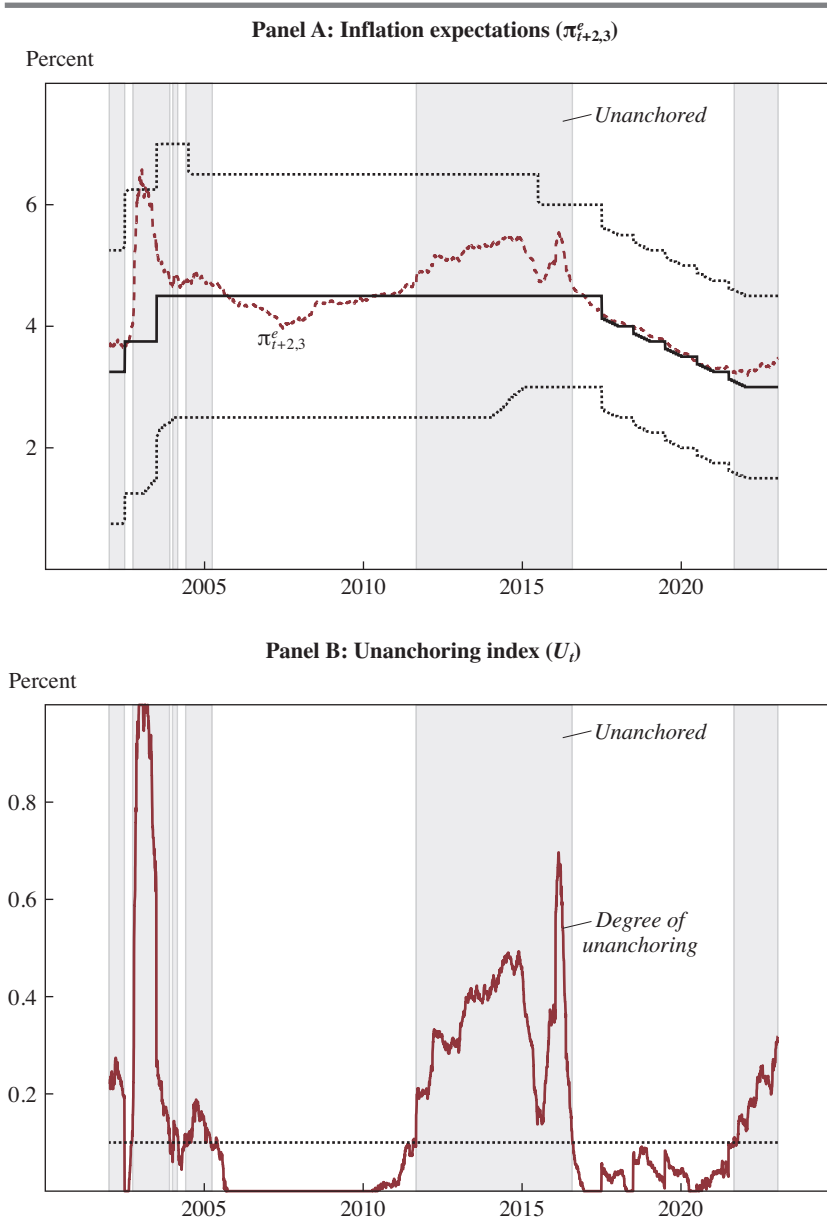
where $\pi^{e}_{t+2,3}$ is the cross-sectional mean of twelve-month inflation expectations between two and three years ahead, $\pi^{*}_{t+2,3}$ is the corresponding inflation target, and $\pi^{max}_{t+2,3}$ is the upper limit of the tolerance band.[6] Inflation expectations are deemed unanchored whenever $U_t > 0.1$.[7] Panel A in figure 2 shows inflation expectations ($\pi^{e}_{t+2,3}$), target, and tolerance band since 2002. Panel B shows the measure of unanchoring. It points to three clear instances of unanchoring of expectations, which we indicate with the shaded areas: in the early 2000s, around the mid-2010s, and, more recently, since 2021.

5. Banco Central do Brasil, "Minutes of the Monetary Policy Committee—Copom," https://www.bcb.gov.br/en/publications/copomminutes/26102022. Results are essentially unchanged if one uses expectations between three and four years ahead.

6. Results are essentially unchanged if we use median expectations. Mean expectations more quickly reflect movements present in the micro data.

7. The 0.1 cutoff accommodates small deviations of expectations from target that arise after 2017 due to announcements of lower targets. In addition, Abib and others (2023) show that a classification of anchored/unanchored regimes based on such a cutoff perfectly lines up with a narrative approach to defining such regimes.

Figure 2. Inflation Expectations and the Unanchoring Index



Sources: Banco Central do Brasil and authors' calculations.

Note: Panel A reports expectations for twelve-month inflation between two and three years ahead, calculated as in Abib and others (2023). Panel B reports the daily unanchoring index U_t, as described in equation (1). Shaded areas correspond to dates in which $U_t > 0.1$.

II. Two Previous Bouts of Inflation and Disinflation

In this section we review two bouts of inflation and disinflation Brazil experienced prior to the pandemic. We focus on domestic developments only, emphasizing aspects related to fiscal and monetary policies. We acknowledge, however, that global and other domestic factors also played a role in both episodes. For brevity, we focus our descriptions on the dynamics of inflation expectations, nominal exchange rates, and the policy rate, abstracting from other potentially informative indicators, such as government bond yields and inflation compensation (also known as "breakeven inflation").

II.A. Episode 1—2002–2005

The first episode was associated with the first time Brazil's current president, Luiz Inácio Lula da Silva, known as Lula, won a presidential election, in 2002. Historically, Lula and the Workers' Party—to which he has always been affiliated—had spoken frequently in favor of voluntarily defaulting on external debt and "auditing" domestic debt.

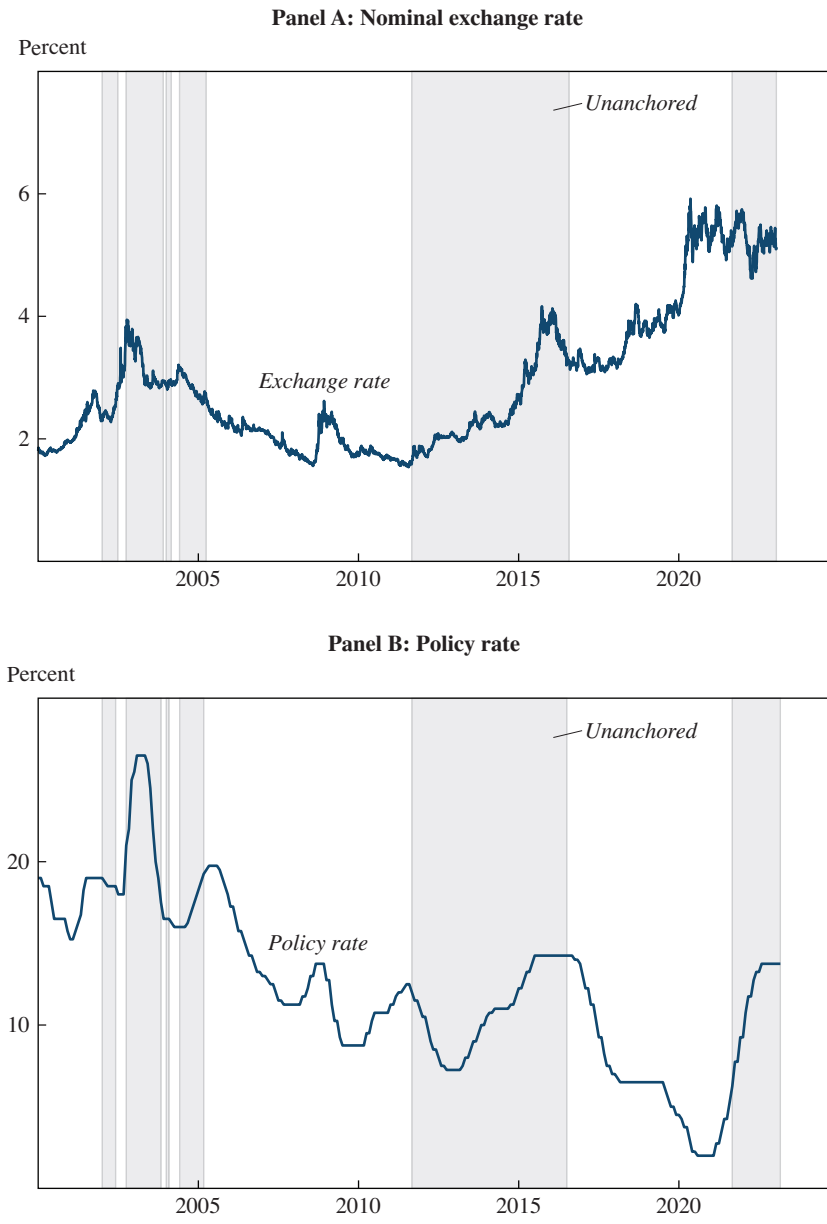
Needless to say, when Lula showed up ahead in the polls over the course of 2002, foreign and local investors dumped Brazilian assets, fearful that existing market-friendly economic policies would be overturned.[8] The currency depreciated dramatically (panel A in figure 3), inflation increased sharply (see figure 1), inflation expectations became unhinged (see figure 2), and the BCB had to tighten policy significantly (panel B in figure 3).[9]

In the first few months of 2002, the BCB had started an easing cycle, after having kept the policy rate constant for about half a year. Between April and September of that year, as concerns about the upcoming election increased, the currency depreciated continuously. Initially, inflation and expectations remained relatively stable. Shortly after Lula won the first election round on October 6 and advanced to a runoff vote to take place later in that month, inflation expectations jumped, inflation picked up, and the currency depreciated sharply.

8. Jensen and Schmith (2005) provide evidence that volatility in the Brazilian stock market increased when election poll results favored Lula. Andrade and Kohlscheen (2010) show evidence that foreign investors became relatively more pessimistic about the prospects for Brazilian risky assets around the 2002 presidential election.

9. In our brief accounts of the high inflation episodes in Brazil, we focus on changes to the policy rate by the BCB. This leaves aside the issue of high real interest rates in Brazil. For discussions pertaining to the level of policy rates, see, for example, Segura-Ubiergo (2012) and Favero and Giavazzi (2005).

Figure 3. Exchange Rate and Policy Rate



Sources: FRED and Banco Central do Brasil.

Note: Panel A shows monthly nominal exchange rate in Brazilian reais per US dollar. Panel B reports the main policy rate (Selic) in percentage points per year. Shaded areas in both panels correspond to the unanchored periods as defined in section I.A.

The BCB responded by raising the policy rate by 300 basis points in an extraordinary meeting on October 14. In its following meetings, the BCB continued to raise the policy rate aggressively, but inflation readings and expectations continued to deteriorate. The concomitant increase of policy rates, inflation, and expectations raised concerns about fiscal dominance. Blanchard (2004) argues that Brazil flirted with a form of fiscal dominance during this period, whereby higher interest rates raised fears of default and led to a depreciation of the Brazilian currency and higher inflation.[10]

Toward the end of 2002, and only after the new government announced a market-friendly economic team (including a new BCB governor), the currency reversed its depreciation trend. In 2003, already under the new government, the BCB continued to increase its policy rate until February. It started a new easing cycle by June 2003, after inflation and expectations had started to decline. The new government announced a series of reforms and fiscal consolidation policies, including higher primary surplus objectives. The adoption of prudent fiscal policies helped calm markets (Werneck 2006; IMF 2007; Paiva Abreu and Werneck 2005).[11]

The whole episode lasted for approximately three years. The BCB increased its policy rate by a total of 850 basis points between October 2002 and February 2003. The economy experienced a brief recession in the first two quarters of 2003, when GDP fell by a total of 1.3 percent.[12] It took until mid-2005 for inflation to return to within the tolerance band and for expectations to reanchor.[13]

II.B. Episode 2—2015–2016

The second episode coincided with the beginning of President Dilma Rousseff's second term, in 2015. Her first term was marked by unsustainable fiscal and quasi-fiscal policies that had been initiated during President Lula's second term. In addition, an abrupt U-turn in monetary policy in 2011, which many commentators associated with political pressure

10. See Loyo (2005) for a discussion.

11. While election years can increase the degree of uncertainty regarding future fiscal and monetary policies, historically they have not necessarily led to unanchoring of expectations. Figure 2 shows that inflation expectations remained well anchored during the 2006 and 2018 elections.

12. Brazilian Business Cycle Dating Committee at FGV IBRE, "CODACE," https://portalibre.fgv.br/en/codace#:~:text=The%20purpose%20of%20the%20Brazilian,business%20cycles%20from%201980%20on.

13. See Bevilaqua, Mesquita, and Minella (2007) for a more detailed account of this episode.

on the BCB, led to unanchoring of inflation expectations (Abib and others 2023).

Following the global financial crisis, still during President Lula's second term, Brazil implemented countercyclical fiscal policies to help alleviate the impact of the ensuing credit crunch. As a result, the country experienced only a mild recession and a strong recovery resumed in the following year.

By mid-2012, the government announced a new set of tax breaks and spending increases to further stimulate the economy. Moreover, the government continued to provide substantial quasi-fiscal stimuli, significantly increasing credit provision through state-owned banks and Brazilian Development Bank (BNDES). This expansion was directly supported by transfers from the Treasury and off-budget measures (Matheson and Pereira 2016). The government also introduced policies aimed at keeping energy and fuel prices artificially low. The large amounts of fiscal and quasi-fiscal support did not help stimulate economic activity (IMF 2013) and resulted in a continuous worsening of the country's fiscal position and the government's credibility (Bloomberg 2016; Spilimbergo and Srinivasan 2019).[14]

Regarding monetary policy, after lowering the policy rate in response to the 2008–2009 credit crunch, the BCB had started a tightening cycle in April 2010, as the economy was growing strongly and inflation edged higher. In the first half of 2011, already under a new governor, the BCB continued to raise rates as inflation moved toward the top of the tolerance band. In its August 31 meeting, however, despite the readings on realized and expected inflation, the BCB surprised markets and cut the policy rate by 50 basis points (Lima 2011). The day after that decision, and in subsequent days, inflation expectations ($\pi^{e}_{t+2,3}$) rose significantly and became unanchored, as shown in figure 2.[15] The BCB doubled down with a sequence of rate cuts, deepening the deterioration of inflation expectations. It was only in April 2013 that the BCB reversed its path and started to raise rates again. By then, inflation had approached 7 percent and inflation expectations were 50 basis points above target.

After being reelected at the end of 2014, President Rousseff started her second term with signs of a possible change in the direction of economic policy. She appointed a market-friendly finance minister, which hinted at a return to a more responsible fiscal policy. She also allowed a reversal of

14. Online appendix A provides a brief account of Brazil's structural fiscal problems since the late 1980s.

15. Abib and others (2023) offer a detailed account of this abrupt monetary policy reversal and provide evidence that it led to the unanchoring of inflation expectations.

policies that had kept energy and fuel prices artificially low, which resulted in a sizable increase in inflation. Despite this increase, prospects of a return to appropriate macroeconomic policies—including responsible fiscal policy and a tighter monetary policy stance—led inflation expectations to improve meaningfully (panel A in figure 2).

The change in fiscal policy, however, did not last. By the second half of 2015, inflation was way above target, expectations had worsened considerably, and the Brazilian real kept depreciating—following a trend that had started in September 2014 (Soto and Cascione 2016). After years of deteriorating fiscal position, the country lost its investment grade and was downgraded to junk status by the end of 2015 (IMF 2015). Once again, fears of unsustainable fiscal policies amid high inflation, unanchored inflation expectations, and tight monetary policy raised the specter of fiscal dominance (Pearson 2016; *Economist* 2016).

A political crisis also ensued, and as time passed, it became increasingly likely Congress would start a process to impeach President Rousseff for running afoul of federal fiscal responsibility laws. Formally, the impeachment process started in December 2015. In May 2016, then vice-president Michel Temer took office. President Rousseff was formally impeached in August 2016 (IMF 2016; Bloomberg 2016).[16]

The prospect of impeachment was associated with the possibility of yet another change in the direction of economic policies. In anticipation of that change, the Brazilian real appreciated and inflation expectations started to move back toward target. President Temer appointed a new economic team, including a new finance minister and a new BCB governor. The new team introduced an ambitious reform agenda and fiscal consolidation policies, which we detail in the next section. The BCB kept a tight monetary policy stance, which also contributed to reanchoring inflation expectations and bringing inflation back to target.

Overall, inflation expectations remained unanchored for five years (between September 2011 and August 2016). Inflation hovered around the top of the tolerance band for most of 2011–2014 and increased significantly above target in 2015, to reach 10.7 percent. Brazilian GDP contracted for eleven consecutive quarters, for a total drop of 8.1 percent.[17]

16. For an account of that period, see Cuevas, Spilimbergo, and Srinivasan (2019).

17. Brazilian Business Cycle Dating Committee at FGV IBRE, "CODACE," https://portalibre.fgv.br/en/codace#:~:text=The%20purpose%20of%20the%20Brazilian,business%20cycles%20from%201980%20on.

Once again, reanchoring and disinflating entailed substantial shifts in economic policy, including the pursuit of sound fiscal policy and significantly tight monetary policy.

III. The Current Disinflation Challenge

III.A. Pre-pandemic Context (2017–2019)

Between 2016 and the start of the pandemic, Brazil underwent a significant effort to address its chronic fiscal challenges through a series of economic reforms and fiscal consolidation measures. While the effects of many of these measures were back-loaded, and hence did little to improve fiscal figures in the short run, they were credible and contributed to an improved perception regarding the course of fiscal policy.

One important reform was the introduction of a cap on public spending by means of a constitutional amendment. With a few exceptions, the cap limited public spending to an amount that increased annually with past inflation.[18] The introduction of such cap stabilized and reduced central government's primary spending.[19] This change was mostly driven by a decline in real public sector compensation and spending items indexed to the minimum wage and cuts to discretionary spending and public investment. Most importantly, the spending cap effectively led society to discuss the trade-offs of any additional public spending, most likely for the first time in the country's history.

Another important measure was the reform of the framework for provision of subsidized and earmarked credit by public institutions. Prior to that reform, public banks and regional development funds were used to provide large amounts of credit at rates that involved substantial subsidies and that were little sensitive to market rates—and hence to monetary policy.[20] The reform created a new set of rates that were linked to the yield curve of government bonds and either eliminated or reduced the implied subsidies. The most emblematic case was the reform of the interest rate on the main

18. The amendment included a few exceptions to this rule. For example, the cap could be violated temporarily in case of calamities, such as a pandemic or war. Moreover, it established a minimum level of spending on education and public health, to prevent these items from being squeezed by other components.

19. See online appendix figure A.1, panel (a).

20. See Bonomo and others (2022) on earmarked credit and monetary policy transmission in Brazil.

funding sources of Brazil's leading national development bank, BNDES.[21] The reform also changed earmarked credit for agriculture and for initiatives aimed at reducing regional inequalities.

Congress also approved a comprehensive pension reform, raising the retirement age, changing the profile of benefits after retirement, and implying significant government savings over time (IMF 2020).[22] Moreover, the country also underwent a divesting program focused on the sale of minority stakes in state-owned enterprises and real estate assets and on efforts to privatize state-owned companies.[23]

III.B. Since the Pandemic

These positive developments, however, were interrupted in early 2020 with the arrival of the COVID-19 pandemic. As in other developed and emerging economies, the pandemic drove a large share of the Brazilian economy to a near complete halt by the second quarter of 2020. In response, large and unprecedented fiscal and monetary policies were quickly put in place. While the responses to the crisis were similar in nature to those implemented elsewhere, policy actions taken by the monetary and fiscal authorities were tailored to Brazil's economic characteristics, its population needs, and the mandates of responding institutions.

Actions by the fiscal authority aimed at low-income households and small and medium-sized firms. Fiscal measures included income transfers, health spending, and loan guarantees, among others (see the online appendix, figure A.2). The federal government implemented one of the largest direct income transfer programs in the world, with an initial disbursement of about 4.5 percent of GDP, reaching more than 60 million people, about a third of the population (McGeever 2021). The sizable fiscal package also included measures to facilitate and subsidize credit to small and medium-sized firms, as well as programs aimed at retaining workers,

21. By 2015, total earmarked credit corresponded to about 50 percent of total credit (about 25 percent of GDP). This large-scale credit provision with subsidized rates hampered the development and deepening of capital markets, in addition to imposing significant costs to the central government budget (Nechio and Serra Fernandes 2022). BNDES accounted for the bulk of earmarked credit to firms. Its subsidized loans, however, were frequently directed to large, unconstrained corporations instead of targeting firms with more difficulties accessing credit through private channels (Bonomo, Brito, and Martins 2015), among other allocation inefficiencies (Carvalho 2013).

22. For details on retirement benefits, see OECD (2021).

23. During that short period, Congress approved other important reforms, including historic labor market reform and the establishment of a credit registry bureau system.

postponing loan payments, and others.[24] The Treasury also played an important role during the crisis by adjusting its bond issuance and repurchasing a record amount of government bonds in moments of distress.

The BCB relied on conventional monetary policy, liquidity provision, and temporary adjustments to its regulatory framework. The monetary authority announced sizable liquidity and credit support programs and adopted measures to temporarily alleviate capital requirements of financial institutions.[25] It also lowered the policy rate to 2 percent, a historically low level (see the online appendix, figure A.2). Finally, the timely and strong response of advanced economies' central banks was also key to restoring market confidence and reversing risk-off attitudes toward emerging economies. As a result of all these unprecedented measures, markets in Brazil stabilized and the financial system was able to withstand the most stressful period of the crisis.

Reflecting the strong response from monetary and fiscal authorities, and the gradual reversal of mobility restrictions, Brazil started to recover by the second half of 2020. The economic rebound was initially concentrated in a few sectors, such as manufacturing and agriculture, and led by consumption of durables. As in other countries, the recovery continued during 2021 and 2022, albeit unevenly due to new waves of COVID-19 and partial shutdowns.

These dynamics were reflected in diverging sectoral price trends, which contributed to the rise of inflation. Similar to other countries, inflationary pressures resulted from supply and demand imbalances brought about by the pandemic, such as supply chain disruptions and the extraordinary monetary and fiscal policies put in place. The relative contributions of these factors varied across countries, as the recent literature has shown (Faria-e-Castro 2021; Cavallo and Kryvtsov 2021; Jordà and Nechio 2023; Hale, Leer, and Nechio 2022). While inflationary pressures were initially concentrated in a few sectors, over time they became more widespread, as revealed by various core inflation measures.

Figure 1 shows the significant rise in headline inflation since mid-2020. It started to reverse its course by mid-2022 as a result of aggressive tax cuts that reduced energy prices, improvements in supply chains, and possibly the first effects of contractionary monetary policy. As figure 2 shows, inflation expectations eventually became unanchored around September 2021.

24. A detailed list of these programs is available on the government's website.

25. Detailed accounts of these and other measures by the BCB are provided in Banco Central do Brasil (2020).

An interesting fact about this high inflation episode is that by the time expectations became unhinged, inflation had already been running high for quite some time. In fact, at that point inflation was close to the peak level it would eventually reach in 2022.[26] Also importantly, by September 2021 the BCB had already done a non-negligible amount of monetary policy tightening, having raised the policy rate from 2 percent in February 2021 to 6.25 percent in September 2021, and market participants expected tightening to continue. Nevertheless, the deterioration of expectations continued, and the BCB further tightened monetary policy by continuously raising the policy rate up to 13.75 percent by August 2022.[27]

The continuing rise of inflation expectations despite inflation having apparently peaked and monetary policy having reached restrictive territory suggests other factors likely played a role in unanchoring expectations. It is noteworthy that heightened fiscal concerns had been present since July 2021, when the government was entertaining new exceptions to the spending cap.[28] As the October 2022 election neared, a range of populist spending measures were unveiled in an attempt by then president Jair Bolsonaro to win votes. During the campaign, the two main candidates, Bolsonaro and Lula, promised considerable fiscal expansion. This contributed to further unanchoring of inflation expectations over the course of 2022.

After the election of Lula da Silva in October 2022 for his third presidential term, the degree of unanchoring started to increase at a faster pace. Since his campaign, but especially since the election, President Lula has been vocal about changing the direction of economic policies. In particular, his government aims to reverse measures and reforms that contributed to fiscal consolidation since 2016. The perception that the country's fiscal position has deteriorated since the election was picked up by several news articles.[29]

26. In August 2021, inflation was at 9.7 percent, whereas the 2022 inflation peak, reached in April 2022, was 12.1 percent.

27. Banco Central do Brasil, "Interest Rate Policy Decisions," https://www.bcb.gov.br/en/monetarypolicy/interestrates.

28. More specifically, a new constitutional amendment to exclude payments due to court-mandated debt from the spending cap was proposed. The proposal was sent to Congress in August and was approved in December 2021.

29. See, for example, Andrade (2022). In addition, around mid-January 2023, President Lula started attacking the BCB's independence and calling for a higher inflation target. Historically, the BCB acted with de facto independence, but it was only in February 2021 that its formal independence was approved by Congress. Lula's attack contributed to further unanchoring of inflation expectations (Harris 2022). For that reason, our sample ends on January 18, 2023.

Table 1. Correlations between Inflation and Gross Debt Projections

	$\rho(\pi^{e}_{t+2,3} - \pi^{*}_{t+2,3}, D_{t+5} - D_{t+1})$		$\rho(\pi^{e}_{t+2,3} - \pi^{*}_{t+2,3}, D_{t+8} - D_{t+1})$	
	Level	*Δ 30-day*	*Level*	*Δ 30-day*
Before unanchored (Dec. 1, 2018 to Aug. 31, 2021)	0.4***	0.03	0.28***	0.09*
Since unanchored (Sept. 1, 2021 to Jan. 18, 2023)	0.68***	0.34***	0.74***	0.28***
Since first round (Oct. 3, 2022 to Jan. 18, 2023)	0.78***	0.68***	0.73***	0.39**

Sources: Banco Central do Brasil and authors' calculations.

Note: Correlations between deviations of inflation expectations from target ($\pi^{e}_{t+2,3} - \pi^{*}_{t+2,3}$) and gross debt/GDP forecast slopes for five ($D_{t+5} - D_{t+1}$) and eight ($D_{t+8} - D_{t+1}$) years ahead.

* 90 percent confidence level, ** 95 percent confidence level, *** 99 percent confidence level, based on Newey-West standard errors

This is the third episode in which unanchoring of inflation expectations seems to be associated with fiscal concerns. Fortunately, this time around we can explore daily expectations data from the Focus Survey for the path of gross public debt (percentage of GDP), available since 2018. Table 1 reports correlations between deviations of inflation expectations from target ($\pi^{e}_{t+2,3} - \pi^{*}_{t+2,3}$) and projected increases in debt/GDP between one and five and one and eight years ahead (i.e., slopes of debt/GDP forecasts).[30] We report correlations for projections in levels and in overlapping thirty-day changes and consider three time periods; prior and since the unanchoring of inflation expectations and since the first election round.

The table shows an increase in the correlation between inflation expectations and debt forecast slopes since unanchoring began, and an additional increase since the election, especially for thirty-day changes in the five-year horizon. This evidence gives credence to our conjecture that unanchoring of inflation expectations and fiscal concerns are once again related in the ongoing episode.

The combination of fiscal concerns and unanchored expectations in a high inflation environment is all too familiar to Brazil. The previous episodes show a range of factors were needed to bring back confidence and reanchor inflation expectations. The BCB had to act strongly. Importantly, actions and expectations of sound fiscal policy were key to restoring market

30. We limit projections to eight years ahead because the number of respondents for the last horizon (nine years) is limited. We work with forecast slopes to eliminate the effects of innovations to debt/GDP and thus isolate perceptions about Brazil's structural fiscal position.

confidence and bringing inflation expectations back to target. In both episodes, disinflation was lengthy and costly in terms of foregone output.

The similarities between the current and the two previous episodes suggest the path ahead for disinflation will be extremely challenging, unless policies change direction.

III.C. Insights for Other Countries?

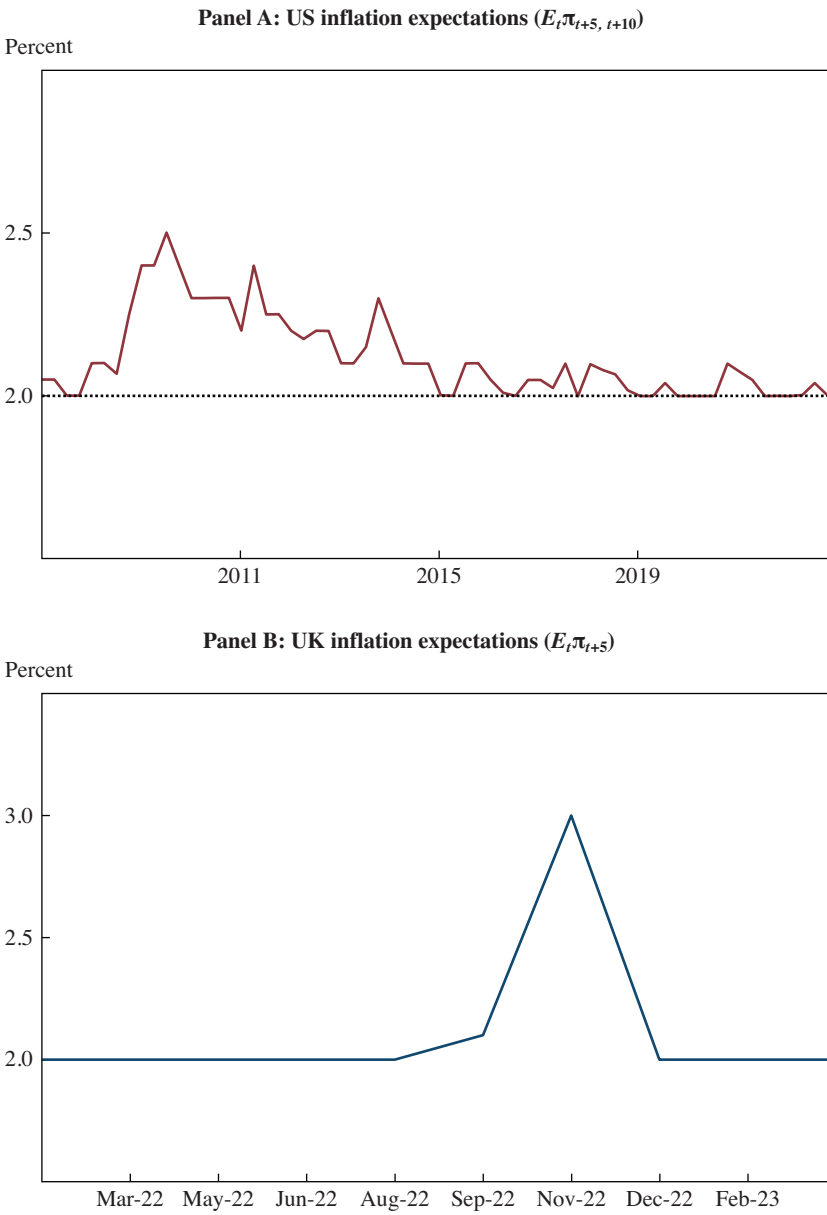
Perhaps the Brazilian experience can help inform the disinflation challenges faced by other countries. Of course, when it comes to cross-country comparisons, making analogies is difficult. It is particularly so when it comes to advanced economies, which have a longer history of low and stable inflation and macroeconomic stability. Nevertheless, two relatively recent episodes in the United States and in the United Kingdom suggest policy credibility can be challenged even in advanced economies.

Panel A of figure 4 shows long-term inflation expectations from US professional forecasters. In particular, we focus on the Survey of Professional Forecasters' median PCE projections for the next five years and five years in the future, a common measure used by the Federal Reserve and market participants to assess inflation expectations (Federal Reserve Board 2022).[31] The series is available from 2007:Q1 to 2023:Q1. The panel shows that long-term inflation expectations in the United States have remained well anchored near the 2 percent inflation target since the pandemic (despite the significant increase in inflation and short-term expectations). Interestingly, the picture also shows a meaningful deviation of long-term inflation expectations from target in the aftermath of the global financial crisis.

Aiming to understand the reasons behind this increase, the Federal Reserve Bank of Philadelphia included a special question in its 2012:Q2 survey, inquiring of participants about the reasons behind their long-term inflation views. In their responses, panelists cited several reasons behind their above-target inflation forecasts. Some worried about the Federal Open Market Committee (FOMC) being too slow to tighten monetary policy at the appropriate time; others worried about the political pressures the FOMC would face, and its willingness to respond to the extraordinarily accommodative fiscal measures introduced in response to the global financial crisis. Finally, participants questioned FOMC members' biases toward

31. The Survey of Professional Forecasters is conducted quarterly by the Federal Reserve Bank of Philadelphia and asks some forty professionals about their views on inflation (among many other variables).

Figure 4. Professionals' Inflation Expectations: United States and United Kingdom



Sources: Federal Reserve Bank of Philadelphia and Bank of England.

Note: Panel A reports inflation expectations from the Survey of Professional Forecasters for five years and five years ahead in the United States. Panel B reports median five-years-ahead inflation expectations from the Bank of England's Market Participants Survey.

higher inflation and reluctance to vote for tighter monetary policy in face of adverse supply shocks.[32]

Turning to the United Kingdom, since 2021 inflation has been persistently elevated and significantly above the Bank of England's 2 percent target. As in other economies, the rise in inflation can be associated with supply and demand imbalances brought about by the pandemic, but also with the energy shock produced by the war in Ukraine (Bank of England 2022). The monetary authority has responded by unwinding its pandemic-related policies and by raising its policy rate from near zero to 4 percent, as of February 2023 (Bank of England 2023).

Against this backdrop, in September 2022 the government announced a fiscal package which, among other measures, lowered (or delayed predicted hikes in) taxes (HM Treasury 2022a). The plan was badly received by markets, which questioned the funding sources and the plan's implications for public debt and inflation (Mackintosh 2022). The weeks that followed saw a sharp increase in bond yields and a depreciation of the British pound. The episode led to the resignation of the prime minister and a nearly complete reversal of the announced policies to "provide confidence in the government's commitment to fiscal discipline" (HM Treasury 2022b, par. 1).

During the process, the Bank of England had to intervene in the bond market due to financial stability concerns. Moreover, as panel B of figure 4 shows, median five-years-ahead inflation expectations from professionals increased from 2 percent to 3 percent in the September and November surveys. Before the announcement and after the policy reversal, long-term expectations were on target. During the time in which inflation expectations rose above target, survey participants revealed concerns with domestic factors as a main driver of their views.[33]

These two episodes and the Brazilian experience share an important component, namely, challenges to policy credibility. While such episodes are somewhat frequent in the case of Brazil, they are rarer when it comes to advanced economies. Nevertheless, the US experience after the global financial crisis and the recent UK episode show central bank credibility can be questioned and long-run inflation expectations can move meaningfully above target. In addition, fiscal concerns were present in both cases, albeit

32. Federal Reserve Bank of Philadelphia, "Long-Run Inflation Forecasts and the FOMC's Longer-Run Goal for Inflation," Survey of Professional Forecasters, https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters.

33. The survey was introduced in February 2022 with the number of participants ranging between thirty and sixty. Results are available on the Bank of England's website.

to different degrees. While we focus on episodes in which inflation expectations moved above target, following the global financial crisis many advanced economies faced risks of downward unanchoring of inflation expectations (Ehrmann 2015). Finally, recent bank failures in the United States have raised the specter of "financial dominance," whereby fears of a broader banking crisis may constrain monetary policy and hamper the Federal Reserve's ability to fight inflation (Summers 2023).

IV. Conclusion

Inflation has risen above target in many countries since the pandemic. In Brazil, together with the rise in inflation, inflation expectations became unanchored. Previous bouts of high inflation in Brazil show policy credibility is frequently challenged, and disinflation is lengthy, economically costly, and requires a combination of factors, which include tight monetary policy and the credible announcement and implementation of sound fiscal policies. Recent developments in the United Kingdom and in the United States show policy credibility can be challenged even in advanced economies.

Empirical evidence from advanced economies available in the literature shows disinflation is usually costly, and more so when policy lacks credibility. While most episodes underlying these conclusions predate the adoption of inflation targeting, the Brazilian episodes we describe in this paper corroborate them. It remains to be seen whether the current disinflation will prove to be as costly as some of the literature for advanced economies and the recent episodes in Brazil suggests.

ACKNOWLEDGMENTS For comments and suggestions, we thank Ken Rogoff (*BPEA* discussant), Maurice Obstfeld, and participants at the Spring 2023 *Brookings Papers on Economic Activity* Conference. Fabián Rivera Reyes provided excellent research assistance. We also benefited from helpful interactions with Marina Perrupato Mendonça. Any errors are our own. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of San Francisco or the Federal Reserve System.

References

Abib, Daniel, João Ayres, Marco Bonomo, Carlos Carvalho, Stefano Eusepi, Silvia Matos, and Marina Perrupato. 2023. "Price Setting When Expectations Are Unanchored." Slides. https://drive.google.com/file/d/1ldGPSw-D2wugZNlKc8_lBg918YGC0-3G/edit.

Andrade, Sandro C., and Emanuel Kohlscheen. 2010. "Pessimistic Foreign Investors and Turmoil in Emerging Markets: The Case of Brazil in 2002." Working Paper 211. Brasilia: Banco Central do Brasil. https://www.bcb.gov.br/pec/wps/ingl/wps211.pdf.

Andrade, Vinicius. 2022. "Morgan Stanley Downgrades Brazil as Lula Imperils Value Play." Bloomberg, November 21. https://www.bloomberg.com/news/articles/2022-11-21/morgan-stanley-cuts-brazil-equities-as-lula-imperils-value-play.

Bacha, Edmar. 2003. "Brazil's Plano Real: A View from the Inside." In *Development Economics and Structuralist Macroeconomics: Essays in Honor of Lance Taylor*, edited by Amitava Krishna Dutt and Jaime Ros. Cheltenham: Edward Elgar Publishing.

Ball, Laurence. 1991. "The Genesis of Inflation and the Costs of Disinflation." *Journal of Money, Credit and Banking* 23, no. 3: 439–52.

Ball, Laurence. 1994. "What Determines the Sacrifice Ratio?" In *Monetary Policy*, edited by N. Gregory Mankiw. Chicago: University of Chicago Press.

Banco Central do Brasil. 2020. *Financial Stability Report*. https://www.bcb.gov.br/content/publications/financialstabilityreport/202010/fsrFullRep.pdf.

Bank of England. 2022. *Financial Policy Summary and Record—December 2022*. https://www.bankofengland.co.uk/financial-policy-summary-and-record/2022/december-2022.

Bank of England. 2023. *Financial Policy Summary and Record—February 2023*. https://www.bankofengland.co.uk/monetary-policy-summary-and-minutes/2023/february-2023#:~:text=The%20Bank%20of%20England's%20Monetary,percentage%20points%2C%20to%204%25.

Bevilaqua, Alfonso S., Mário Mesquita, and André Minella. 2007. "Brazil: Taming Inflation Expectations." Working Paper 129. Brasilia: Banco Central do Brasil. https://www.bcb.gov.br/pec/wps/ingl/wps129.pdf.

Blanchard, Olivier. 2004. "Fiscal Dominance and Inflation Targeting: Lessons from Brazil." Working Paper 10389. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w10389.

Blinder, Alan S. 2023. "Landings, Soft and Hard: The Federal Reserve, 1965–2022." *Journal of Economic Perspectives* 37, no. 1: 101–20.

Bloomberg. 2016. "Dilma Rousseff Learns Fiscal Crime Doesn't Pay." Bloomberg, August 31. https://www.bloomberg.com/opinion/articles/2016-08-31/dilma-rousseff-learns-fiscal-crime-doesn-t-pay.

Bonomo, Marco, Ricardo D. Brito, and Bruno Martins. 2015. "The After Crisis Government-Driven Credit Expansion in Brazil: A Firm Level Analysis." *Journal of International Money and Finance* 55:111–34.

Bonomo, Marco, Bruno Martins, Bruno Perdigão, and Carlos Carvalho. 2022. "Crédito direcionado e seus efeitos sobre a transmissão da política monetária." In *Para não esquecer: Políticas públicas que empobrecem o Brasil*, edited by Marcos Mendes. Rio de Janeiro: Autografia.

Carvalho, Daniel. 2013. "The Real Effects of Government-Owned Banks: Evidence from an Emerging Market." *Journal of Finance* 69, no. 2: 577–609.

Cavallo, Alberto, and Oleksiy Kryvtsov. 2021. "What Can Stockouts Tell Us about Inflation? Evidence from Online Micro Data." Working Paper 29209. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29209.

Cecchetti, Stephen G., Michael E. Feroli, Peter Hooper, Frederic S. Mishkin, and Kermit L. Schoenholtz. 2023. "Managing Disinflations." Working paper. https://www.chicagobooth.edu/-/media/research/igm/docs/usmpf-2023-conference-version.pdf.

Cecchetti, Stephen G., and Stefan Krause. 2002. "Central Bank Structure, Policy Efficiency, and Macroeconomic Performance: Exploring Empirical Relationships." *Federal Reserve Bank of St. Louis Review* 84, no. 4.

Cuevas, Alfredo, Antonio Spilimbergo, and Krishna Srinivasan. 2019. "Brazil: Economic Policies and Institutions in the Twenty-First Century." *Journal of Development Perspectives* 3, nos. 1–2: 96–110.

Economist. 2016. "Irredeemable?" January 2. https://www.economist.com/briefing/2016/01/02/irredeemable.

Ehrmann, Michael. 2015. "Targeting Inflation from Below: How Do Inflation Expectations Behave?" *International Journal of Central Banking* 11, no. S1: 213–49.

Faria-e-Castro, Miguel. 2021. "Fiscal Policy during a Pandemic." *Journal of Economic Dynamics and Control* 125:104088.

Favero, Carlo A., and Francesco Giavazzi. 2005. "Inflation Targeting and Debt: Lessons from Brazil." In *Inflation Targeting, Debt, and the Brazilian Experience, 1999 to 2003*, edited by Francesco Giavazzi, Ilan Goldfajn, and Santiago Herrera. Cambridge, Mass.: MIT Press.

Federal Reserve Board. 2022. "Statement on Longer-Run Goals and Monetary Policy Strategy." https://www.federalreserve.gov/monetarypolicy/files/fomc_longerrungoals.pdf.

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton, N.J.: Princeton University Press.

Ha, Jongrim, M. Ayhan Kose, and Franziska Ohnsorge. 2019. *Inflation in Emerging and Developing Economies: Evolution, Drivers, and Policies*. Washington: World Bank Group.

Hale, Galina, John Leer, and Fernanda Nechio. 2022. "Inflationary Effects of Fiscal Support to Households and Firms." Working Paper 2023-02. Federal Reserve Bank of San Francisco. https://www.frbsf.org/wp-content/uploads/sites/4/wp2023-02.pdf.

Harris, Bryan. 2022. "Lula's Attacks on Brazil Central Bank Alarm Investors." *Financial Times*. https://www.ft.com/content/011951a9-9db2-4834-87d9-f97df82b38c5.

HM Treasury. 2022a. "Chancellor Announces New Growth Plan with Biggest Package of Tax Cuts in Generations." Gov.UK, September 23. https://www.gov.uk/government/news/chancellor-announces-new-growth-plan-with-biggest-package-of-tax-cuts-in-generations.

HM Treasury. 2022b. "Chancellor Brings Forward Further Medium-Term Fiscal Plan Measures." Gov.UK, October 17. https://www.gov.uk/government/news/chancellor-brings-forward-further-medium-term-fiscal-plan-measures.

International Monetary Fund (IMF). 2007. "Brazil: Helping Calm Financial Markets." https://www.imf.org/external/np/exr/articles/2007/112107.htm.

International Monetary Fund (IMF). 2013. *Brazil: Staff Report for the 2013 Article IV Consultation*. Washington: IMF. https://www.imf.org/external/pubs/ft/scr/2013/cr13312.pdf.

International Monetary Fund (IMF). 2015. *Brazil: 2014 Article IV Consultation—Staff Report; Press Release; and Statement by the Executive Director for Brazil*. Washington: IMF. https://www.imf.org/external/pubs/ft/scr/2015/cr15121.pdf.

International Monetary Fund (IMF). 2016. *Brazil: 2016 Article IV Consultation—Press Release; Staff Report; and Statement by the Executive Director for Brazil*. Washington: IMF. https://www.imf.org/external/pubs/ft/scr/2016/cr16348.pdf.

International Monetary Fund (IMF). 2020. *Brazil: 2020 Article IV Consultation—Press Release; Staff Report; and Statement by the Executive Director for Brazil*. Washington: IMF. https://www.imf.org/en/Publications/CR/Issues/2020/12/01/Brazil-2020-Article-IV-Consultation-Press-Release-Staff-Report-and-Statement-by-the-49927.

Jensen, Nathan M., and Scott Schmith. 2005. "Market Responses to Politics: The Rise of Lula and the Decline of the Brazilian Stock Market." *Comparative Political Studies* 38, no. 10: 1245–70.

Jordà, Òscar, and Fernanda Nechio. 2023. "Inflation and Wage Growth since the Pandemic." Working Paper 2022-17. Federal Reserve Bank of San Francisco. https://www.frbsf.org/economic-research/publications/working-papers/2022/17/.

Lima, Mario Sergio. 2011. "Brazil Job Creation Rose More Than Expected in September." Bloomberg, October 18. https://www.bloomberg.com/news/articles/2011-10-18/brazil-job-creation-rose-more-than-expected-in-september-1-#xj4y7vzkg.

Loyo, Eduardo. 2005. Comment on "Fiscal Dominance and Inflation Targeting: Lessons from Brazil" by Olivier Blanchard. In *Inflation Targeting, Debt, and the Brazilian Experience, 1999 to 2003*, edited by Francesco Giavazzi, Ilan Goldfajn, and Santiago Herrera. Cambridge, Mass.: MIT Press.

Mackintosh, James. 2022. "Britain's Financial Disaster Is a Warning to the World." *Wall Street Journal*, October 1. https://www.wsj.com/articles/britains-financial-disaster-is-a-warning-to-the-world-11664596851.

Matheson, Troy D., and Joana Pereira. 2016. "Fiscal Multipliers for Brazil." Working Paper 2016/079. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Fiscal-Multipliers-for-Brazil-43831.

McGeever, Jamie. 2021. "Brazil GDP Drops 4.1% in 2020, COVID-19 Surge Erodes Rebound." Reuters, March 3. https://www.reuters.com/article/us-brazil-economy-gdp/brazil-gdp-drops-4-1-in-2020-covid-19-surge-erodes-rebound-idUSKBN2AV1FZ.

Nechio, Fernanda, and Bruno Serra Fernandes. 2022. "Brazil: Covid-19 and the Road to Recovery." In *The Monetary-Fiscal Policy Nexus in the Wake of the Pandemic*, no. 122 *BIS Papers*. Basel: Bank for International Settlements.

OECD. 2021. *Pensions at a Glance 2021: OECD and G20 Indicators*. Paris: OECD Publishing.

Paiva Abreu, Marcelo de, and Rogerio Ladeira Furquim Werneck. 2005. "The Brazilian Economy from Cardoso to Lula: An Interim View." Texto Para Discussão 504. Rio de Janeiro: Pontifical Catholic University of Rio de Janeiro.

Pearson, S. 2016. "Surging Inflation Leaves Brazil's Rate Setters with Headache." *Financial Times*, January 18. https://www.ft.com/content/8d67130c-bdfb-11e5-a8c6-deeeb63d6d4b.

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.

Sargent, Thomas J. 1982. "The Ends of Four Big Inflations." In *Inflation: Causes and Effects*, edited by Robert E. Hall. Chicago: University of Chicago Press.

Sargent, Thomas J. 2013. "Stopping Moderate Inflations: The Methods of Poincaré and Thatcher." In *Rational Expectations and Inflation*, edited by Thomas J. Sargent. Princeton, N.J.: Princeton University Press.

Segura-Ubiergo, Alex. 2012. "The Puzzle of Brazil's High Interest Rates." Working Paper 2012/062. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2016/12/31/The-Puzzle-of-Brazil-s-High-Interest-Rates-25751.

Soto, Alonso, and Silvio Cascione. 2016. "Brazil Inflation Ends 2015 above Target, Cenbank Vows Action." Reuters, January 8. https://www.reuters.com/article/us-brazil-economy-inflation/brazil-inflation-ends-2015-above-target-cenbank-vows-action-idINKBN0UM2IO20160108.

Spilimbergo, Antonio, and Krishna Srinivasan. 2019. *Brazil: Boom, Bust, and the Road to Recovery*. Washington: International Monetary Fund.

Summers, Lawrence H. 2023. "Lessons from the SVB Bank Failure." Video. https://bcf.princeton.edu/events/lawrence-h-summers-on-lessons-from-the-svb-bank-failure/.

Tetlow, Robert J. 2022. "How Large Is the Output Cost of Disinflation?" Finance and Economics Discussion Series. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econres/feds/files/2022079pap.pdf.

Werneck, Rogério L. F. 2006. "An Evaluation of the 2003 Tax Reform Effort in Brazil." *Brazilian Journal of Political Economy* 26, no. 1: 75–94.

ESWAR PRASAD
Cornell University

Has China's Growth Gone from Miracle to Malady?

ABSTRACT China's remarkable run of persistently high growth in recent decades is all the more stunning in light of the country's low levels of financial and institutional development, state-dominated economy, and nondemocratic government. Notwithstanding the inefficient and risky growth model, the government has maneuvered the economy around various stresses without any major financial or economic crash. With a shrinking labor force and declining efficiency of investment, raising productivity growth is key to maintaining reasonable GDP growth. Unbalanced reforms, a schizophrenic approach to the role of the market versus the state, and strains in financial and property markets could result in significant volatility but a financial or economic collapse is not in the cards.

Several unprecedented economic, financial, and political events have occurred in recent decades, many of them short-lived and most of them unfavorable in some form or another. Then there is China's growth over the last few decades, which stands out as a positive historical (and persistent) anomaly by any measure. Mean reversion, the middle income trap, or any number of formal, informal, and statistical models have for many years now been used to predict an imminent sharp decline, if not collapse, in China's growth. China's low levels of financial and institutional development, state-dominated economy, nondemocratic government, and manifold other inadequacies should have dragged down growth. And yet,

Conflict of Interest Disclosure: Eswar Prasad is the Tolani Senior Professor of Trade Policy and professor of economics at Cornell University, and a senior fellow at the Brookings Institution. Brookings is financed through the support of a diverse array of foundations, corporations, governments, and individuals, as well as an endowment. A list of donors can be found in our annual reports published online at https://www.brookings.edu/about-us/annual-report/. The findings, interpretations, and conclusions in this paper are solely those of its author and are not influenced by any donation.

Brookings Papers on Economic Activity, Spring 2023: 243–270 © 2024 The Brookings Institution.

until the COVID-19 pandemic rocked it back on its heels, the Chinese economy powered through periods of domestic and global turmoil seemingly unscathed. Is the much anticipated and long-foretold day of reckoning finally at hand?

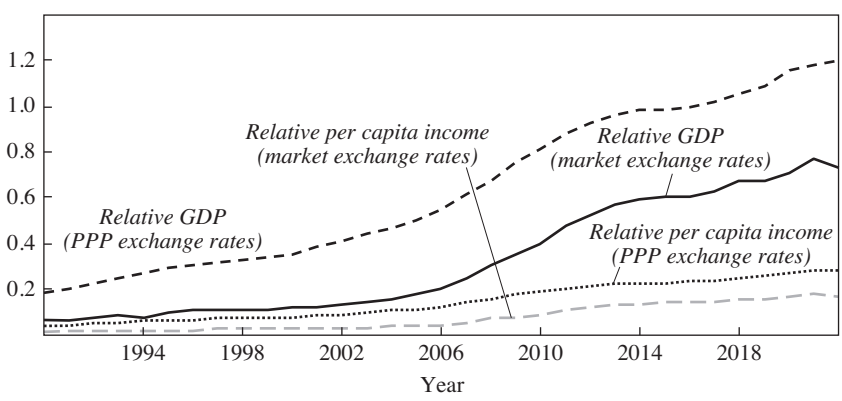
China has had a remarkable run of growth, lofting it from low income status three decades ago to upper middle income status now. China's GDP was $18.3 trillion in 2022 (at market exchange rates), 73 percent that of the United States and ten times more than the comparable ratio of 7 percent in 1990 (figure 1). The gap in per capita incomes is larger. China's per capita income is now roughly $13,000, approximately 17 percent that of the United States, compared to less than 2 percent in 1990. In purchasing power parity (PPP) exchange rate terms, China's economy is now 1.2 times the size of the United States. Over the period 2007–2022, China accounted for 35 percent of global nominal GDP growth (at market exchange rates), compared to 27 percent for the United States (IMF 2022).[1]

China's GDP growth in the last three decades has been reliant on investment growth financed by an inefficient banking system and has occurred against the backdrop of a bloated state sector, authoritarian government, and weak institutional framework. A striking aspect of China's economic performance has been the government's ability to manage the severe economic and financial stresses that have built up as a result of the highly inefficient and risky growth model it has embraced. The government has maneuvered the economy around the seemingly inevitable prospects, at various points, of a banking crisis, massive currency devaluation, housing market meltdown, and economic collapse. Notwithstanding numerous ructions in financial and housing markets and in GDP growth, these plausible doomsday scenarios have yet to materialize.

Yet, each of these near misses has exacted a toll. The huge buildup in domestic debt, the loss of $1 trillion in reserves during 2015–2016, and the highly volatile prices of stocks, property, and other assets are emblematic of the challenges the economy has had to contend with (Setser 2018).

1. The growth contribution figures are based on nominal GDP in US dollars measured at market exchange rates. Measured in PPP exchange rates, China and the United States accounted for 26 percent and 13 percent, respectively, of global GDP growth over this period. While China's GDP is larger in PPP terms, over the period 2007–2022 the renminbi appreciated relative to the US dollar by 12 percent at market exchange rates but *depreciated* by 34 percent at PPP exchange rates (end-of-year exchange rates). China's nominal GDP in US dollars at market exchange rates was 5.1 times larger in 2022 than in 2007. At PPP exchange rates, China's 2022 GDP was 3.4 times its 2007 level. When measured using GDP at constant prices and PPP exchange rates, China accounts for 36 percent of global growth from 2007 to 2022.

Figure 1. Relative GDP and Per Capita Income: China and the United States



Source: IMF *World Economic Outlook*, October 2022.
Note: This chart shows China's GDP and per capita GDP relative to the United States based on market and PPP exchange rates.

They also highlight erosions in the efficacy of macroeconomic policy tools and the government's shrinking capacity to respond effectively to adverse domestic and external circumstances.

Beijing now faces a number of policy dilemmas, some of which are of its own creation—how to continue deleveraging while maintaining growth; how to reduce energy-intensive production while the economy continues to rely on heavy industry; how to get markets to exert financial discipline even as Beijing tries to strengthen state control; how to restrain wealth inequality while relying on the private sector to generate more wealth; and how to encourage private sector innovation while cutting successful private enterprises down to size.

Beijing's attempts to resolve these inherently contradictory impulses in the guise of market-oriented socialism will inevitably lead to further stumbles and accidents. The government's policy approach, while driven by the right objectives, could generate more uncertainty and volatility in the short run, which in turn could reduce public support for much needed reforms to bolster long-term productivity and growth.

Even if no crises materialize and the government succeeds in steering the economy through murky waters, unfavorable demographics, high debt levels, and an inefficient financial system will increasingly constrain China's growth. This gloomy picture has to be offset by the recognition that, for an economy that now accounts for nearly one-fifth of global GDP

at market exchange rates (IMF 2022), even a more restrained growth rate relative to its own historical standards will mean the continuation of an extraordinary growth streak.

I. Sources and Structure of Growth

China's official GDP data have become the focus of much attention and some controversy. Whether China's GDP is correctly measured remains a subject of debate, with some even arguing that growth might be understated in certain periods. By and large, scholars seem to agree that data on annual growth rates have become reasonably reliable in recent years, matching up with various other indicators of economic activity (while quarterly growth rates often seem too smooth to be realistic).[2]

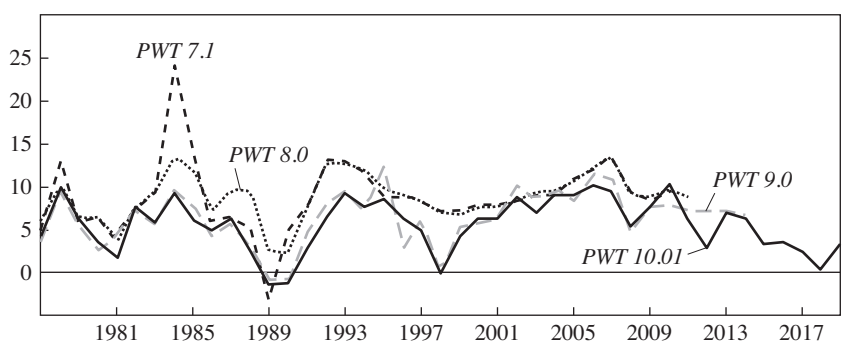
Among other sources of alternative data, the Penn World Table (PWT) is used widely in macroeconomic research since it provides a breakdown of GDP into different expenditure components, in both local currency and international prices, and on a consistent basis for a large number of countries (Feenstra, Inklaar, and Timmer 2015). Based on the then available PWT version 8.0 data, Pritchett and Summers (2014) highlight China's phenomenal run of per capita real GDP growth.[3] They note that "China's experience from 1977 to 2010 already holds the distinction of being the only instance, quite possibly in the history of mankind, but certainly in the data, with a sustained episode of super-rapid (> 6 ppa) growth for more than 32 years" (Pritchett and Summers 2014, 36). They point to Brazil, Japan, and Southeast Asian economies as having experienced periods of sustained high growth before shifting to long periods of low to zero growth, with China being a huge statistical outlier by any measure. Based on analysis of cross-country historical data, they make a strong statistical case for mean reversion pulling down China's growth.

In the latest version of the PWT (10.01), China's growth looks a little less remarkable. Figure 2 shows a comparison of China's growth based on

2. Chen and others (2019) argue that China's official statistics overstate GDP growth by about 1.8 percentage points per annum over the period 2010–2016. Clark, Pinkovskiy, and Sala-i-Martin (2020) reach the opposite conclusion for a longer sample, while Holz (2014) and Fernald, Hsu, and Spiegel (2021) conclude that the official data are mostly reasonable, if a tad too smooth.

3. In later sections of their paper, Pritchett and Summers (2014) use PWT 7.1 to project China's dollar GDP into the future in order to build on some earlier calculations using that version.

Figure 2. Per Capita Real GDP Growth across Different PWT Versions



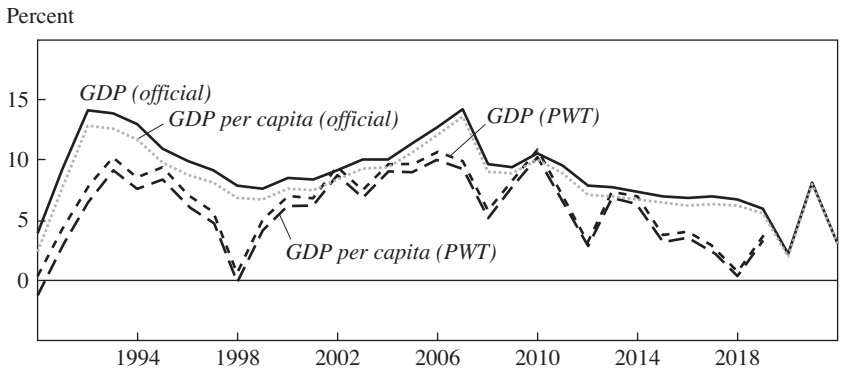
Sources: Penn World Table versions 7.1, 8.0, 9.0, and 10.01.

Note: Real GDP at constant 2005 national prices in PWT 7.1 and 8.0 data, real GDP at constant 2011 national prices in PWT 9.0 data, and real GDP at constant 2017 national prices in PWT 10.01 data. The specific variable used in these calculations is $RGDP^{NA}$, which is the best measure for evaluating the "growth performance of economies" (Feenstra, Inklar, and Timmer 2015).

data from different versions of the PWT. Average annual per capita GDP growth over 1977–2010—the sample in Pritchett and Summers (2014)—is 8.75 based on PWT 8.0 data and roughly 6 percent based on PWT 9.0 and 10.01 (which has data through 2019). China's per capita GDP growth rate was 3.9 percent from 2011 through 2019 per PWT 10.01 and 6 percent from 2011 through 2022 in the official data. Irrespective of which data source (and which version of that source) one chooses to use, Pritchett and Summers's substantive point stands—that China's growth over the last five decades has been spectacular and unique in recent history. And their point that China's growth would perforce slow down seems to be holding up as well, although average growth of 6 percent over a decade is slow only by China's own historical standards.

Figure 3 shows China's growth over the last three decades based on official data as well as data from the PWT 10.01. The trend decline in growth over the last decade and a half is apparent in the data. The pandemic-induced slowdown in 2020 was followed by a strong rebound in 2021, before the government's draconian COVID-19 lockdown policies pulled growth back down to 3 percent in 2022. In late 2022, the Chinese government's reversal of its increasingly untenable zero-COVID-19 policy briefly lifted domestic stock markets and buoyed optimism about China's short-term growth prospects. The government's volte-face can be taken as a sign

Figure 3. China: Annual Growth in Real GDP and Per Capita Real GDP



Sources: World Bank's World Development Indicators and Penn World Table 10.01.

Note: Real GDP in constant local currency units in World Bank data and real GDP at constant 2017 national prices in PWT data. The specific PWT variable used in these calculations is $RGDP^{NA}$, which is the best measure for evaluating the "growth performance of economies" (Feenstra, Inklar, and Timmer 2015).

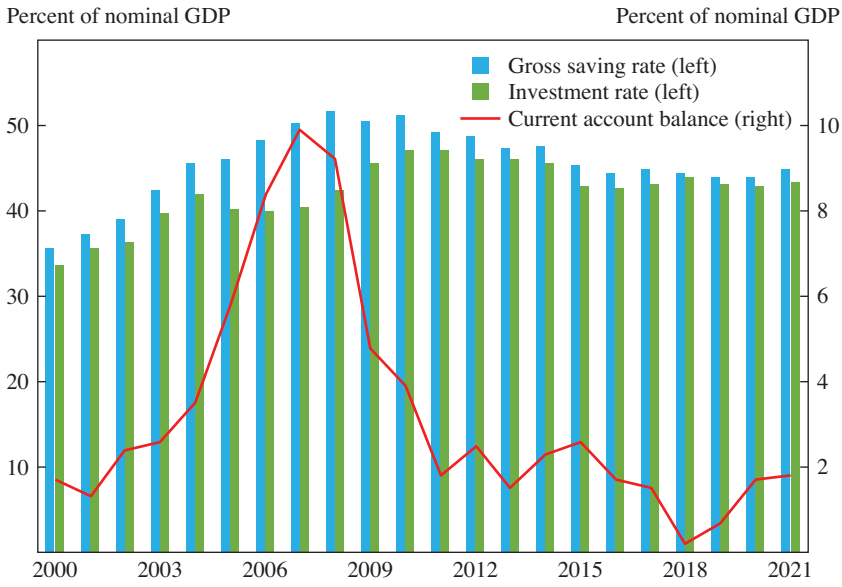
of flexibility or, less favorably, its dogged unwillingness to acknowledge policy errors and undertake a course correction until left with no choice. In any event, the economy seems set for a stronger performance during 2023 than in 2022, but growth beyond that will depend on how structural changes that have been underway in the economy play out.

I.A. Fixing Imbalances

For many years, the Chinese government has touted its goals of rebalancing the economy. This is taken to mean reducing the reliance on investment-heavy growth and getting household consumption to be the key contributor to GDP growth; generating more growth from the services sector rather than low-skill, low-wage manufacturing; and shifting away from capital-intensive growth in a manner that improves employment growth. These could be seen as different facets of the same objective.

In the decade between the Asian and global financial crises, China's large and rising current account surplus—which peaked at 10 percent of GDP in 2007—was seen as an important symptom of domestic imbalances (and as a major contributor to global imbalances). High investment and even higher saving were anomalies to be fixed. Figure 4 shows that some of this rebalancing did take place, with China's current account surplus falling to and remaining at around 2 percent of GDP since 2011. Both saving

Figure 4. Saving, Investment, and Current Account Balance



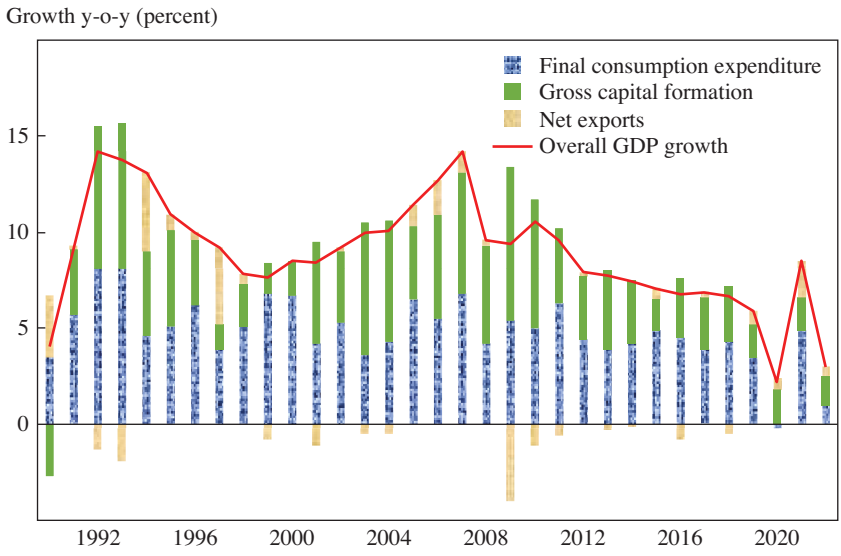
Source: CEIC.

and investment have fallen from their peaks but remain high by international standards. For instance, the investment-to-GDP ratio has declined by 4 percentage points from its 2010–2011 peak but still remains high at 43 percent.

Did the external rebalancing reflect internal rebalancing? Figure 5 shows the contributions of different expenditure components to real GDP growth. Domestic rebalancing initially lagged behind external rebalancing. For much of the last two decades, investment has been the key contributor to growth. This pattern intensified in the immediate aftermath of the global financial crisis, with investment growth accounting for about two-thirds of GDP growth during 2009–2010. Since then, household consumption has become the main contributor to growth. In 2020, with household consumption collapsing during the pandemic, investment again dominated growth. This pattern was reversed in 2021 as pent-up demand sent household consumption surging. It is worth noting that net exports have in general been a relatively modest contributor to China's overall growth.

Given that China has a capital-to-labor ratio that is much lower than that of advanced economies, one could argue that more rather than less

Figure 5. Growth Contributions of Demand Components



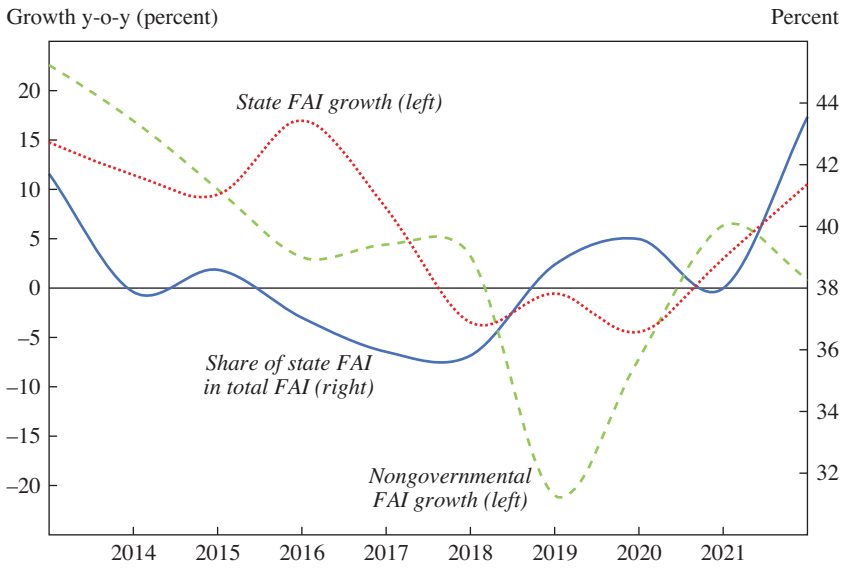
Sources: CEIC and National Bureau of Statistics, China.

investment is desirable.[4] For instance, China's capital-to-labor ratio is only about 28 percent that of the United States. However, recent investment has been driven by the public (state) sector rather than the nongovernmental sector. In 2022, for instance, state investment amounted to 44 percent of total fixed asset investment, a significant increase relative to the corresponding ratio of about 36 percent during 2017–2018 (figure 6).[5] This is not inherently a problem—after all, investment in private sector firms, especially smaller ones, rather than large, state-owned enterprises can be

4. The ratios mentioned in this paragraph are based on PWT 10.01 estimates of capital stocks, total factor productivity (TFP), and GDP at constant 2017 national prices and employment levels for the two countries in 2019. Estimates of capital stocks and GDP at current PPPs yield similar ratios. In principle, estimates of the labor force rather than employment levels should be used in such calculations, but this would not affect the ratio much given current estimates of unemployment rates in both countries.

5. The breakdown of investment into state and nongovernmental is based on fixed asset investment data. This is different from the national income accounts concept of gross fixed capital formation (which is the measure used earlier in this paper). A sectoral breakdown of gross fixed capital formation is not available. For a critical evaluation of various measures of investment in China, see Lardy (2020).

Figure 6. Fixed Asset Investment



Sources: CEIC and National Bureau of Statistics, China.

Note: Growth rates are year-over-year based on monthly data. Starting in 2018, the National Bureau of Statistics stopped providing data on levels of fixed asset investments (FAI). The ratio of state FAI to total FAI is based on imputed end-of-year data.

much riskier. But in China state-owned enterprises, which have collectively received a disproportionate share of bank credit, have typically not generated strong returns on those investments.[6]

The recent collapse in nongovernmental investment growth (figure 6), with state investment accounting for nearly all of the growth in overall fixed asset investment in 2022, is a sign that private businesses might be wary of increasing investment when they see the economic and political environments as unfavorable. Moreover, China's capital-to-*output* ratio is in fact about 50 percent higher than that of the United States.[7] This reflects lower levels of total factor productivity (TFP) and human capital in China

6. As one example of the differentials in productivity and profitability between state-owned and private enterprises, Jurzyk and Ruane (2021) find that publicly listed state-owned enterprises are less productive and profitable than publicly listed firms in which the state has no ownership stake.

7. The World Bank, "GDP Growth (Annual %)—United States, China," https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=US-CN.

relative to the United States. This implies that increasing investment might not be the optimal way to generate growth (Rogoff 2022).[8]

Another aspect of rebalancing that has played out over the last three decades is the gradual increase in the services sector's share of aggregate GDP as well as employment (figure 7). By 2012, the services sector accounted for the largest shares of output and employment and, by 2021, this sector accounted for more than 50 percent of GDP and close to 50 percent of aggregate employment. With this sector accounting for a substantial fraction of employment growth in recent decades, the availability of financing and productivity growth in this sector will be important determinants of China's growth.[9]

In short, while the trajectory has been uneven, there has been significant progress toward the objective of growth rebalancing, with household consumption becoming the key driver of growth and the services sector becoming more prominent than manufacturing.

I.B. Prospects

Speculating about China's medium- and long-term growth prospects has been a growth industry in itself—one that has enormous implications for the world economy as well.[10] Pritchett and Summers (2014) and many other authors have pointed to the risks of extrapolating China's future growth based on its performance over the last few decades. There is a large margin of uncertainty surrounding any projections of growth in an economy undergoing substantial structural change. One can at best use the growth of various factors that go into the aggregate production function as an indicator of what the future might hold.

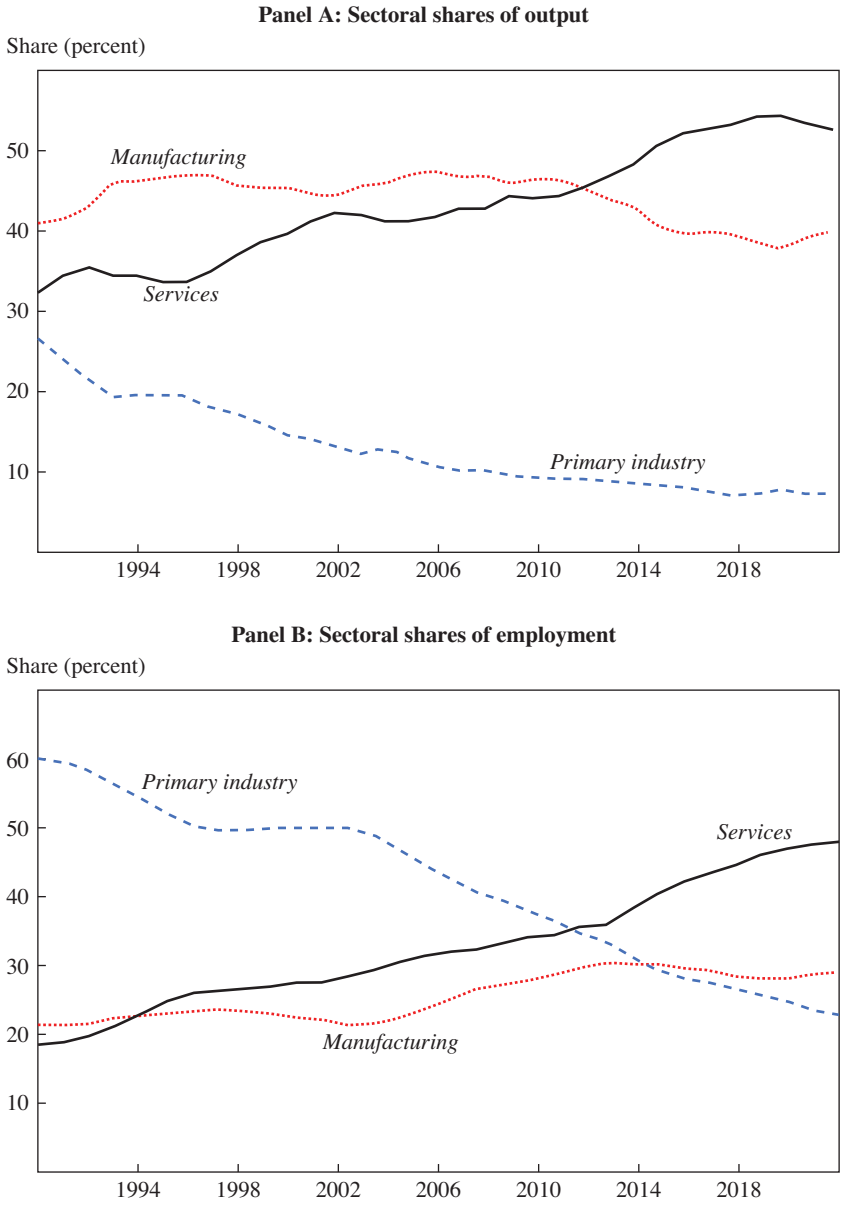
China's labor force, defined as the population in the 15–64 age range, is shrinking (figure 8). By 2030, it is expected to decline at the rate of about 1 percent per year, acting as a drag on growth (immigration is minimal and unlikely to offset this decline). Of course it is not just the number of bodies but the skills embedded in the workforce that matter for output. Is the quality

8. For more analysis of these shifts in investment patterns and their implications, see Lardy (2014, 2019).

9. Over the two decades from 1991 to 2021, average annual employment growth in industry and services was 1.5 percent and 3.6 percent, respectively. Overall employment growth averaged only 0.5 percent over this period, as employment in the primary sector continued to shrink. The official urban unemployment rate was 4 percent at the end of 2021.

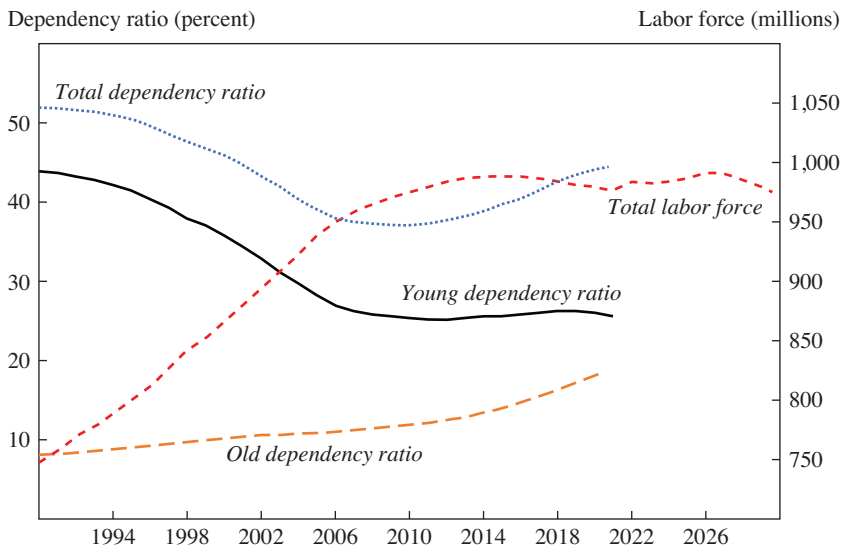
10. For some recent examples, see Bailliu and others (2017); Dollar, Huang, and Yao (2020); Rosen (2021); and Benzell and others (2022). For discussion of the middle income trap and potential implications for China's growth, see Eichengreen, Park, and Shin (2012) and Kharas and Gill (2020).

Figure 7. Changes in the Structure of the Chinese Economy



Sources: CEIC Data and National Bureau of Statistics, China.

Figure 8. Demographics and Labor Force



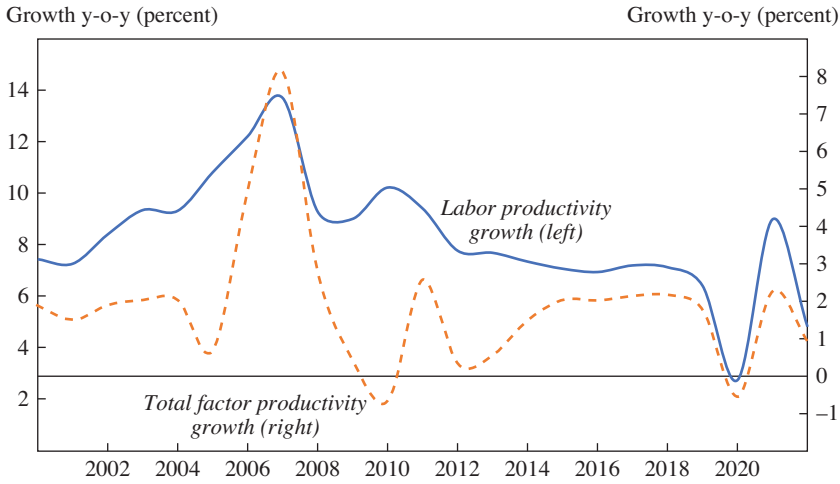
Sources: United Nations; World Bank; and National Bureau of Statistics, China.
Note: Labor force forecasts for 2023–2030 are from the United Nations.

of the labor force improving? China's human capital index in the PWT, based on years of schooling and the returns to education, shows an increase from 1.96 in 1990 to 2.44 in 2010 and then to 2.69 in 2019 (comparable figures for the United States are 3.44, 3.70, and 3.75, respectively).[11] The increase in human capital embedded in each unit of labor could partially offset shrinkage in the labor force, although part of the increase in the PWT's measure of human capital could itself simply reflect higher returns to education on account of a rising capital-labor ratio.

Higher investment growth could pick up some of the slack but that has many risks, as discussed earlier. That leaves productivity: for all the inefficiencies that seem to pervade its economy, China has generated decent TFP growth on average over the last few decades.[12] But as figure 9 shows, productivity growth has been muted over the last decade, with annual TFP growth averaging only about 1 percent per annum. Clearly, China's growth will run aground without an improvement in productivity growth. The government seems to recognize this, but before we discuss its policy approach,

11. These numbers are taken from PWT 10.01 and refer to the variable *hc*.
12. See Bosworth and Collins (2008) and Sasaki and others (2021).

Figure 9. Growth in Labor Productivity and Total Factor Productivity



Sources: CEIC and the Conference Board.

we must consider the possibility that the economy could come off the rails, validating warnings about all the overt and hidden imbalances that have been building up for many years.

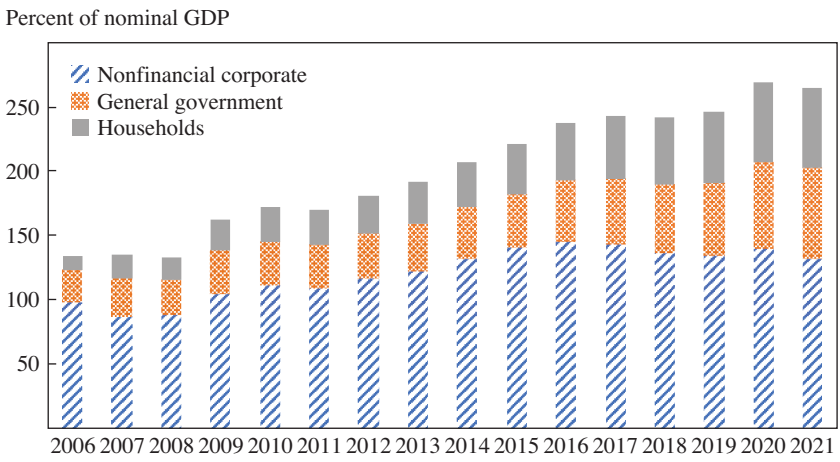
II. Vulnerabilities

A number of domestic and external financial vulnerabilities loom in the background, raising the specter of China's economy being vulnerable to a crash similar to that experienced by other high-flying economies, such as the Four Asian Tigers. There are grounds for concern, but at least some of the traditional red flags—such as high levels of foreign currency–denominated external debt, which have caused numerous emerging market economy crashes—are not present in China's case.

II.A. Debt Distress?

China's overall debt levels and the possibility of a financial crash that spills over to global financial markets have been a significant concern for many years now. Debt levels have indeed risen over time, although gross debt levels are not substantially out of line with other major economies such as the United States and Japan. One important difference is that, as a ratio to nominal GDP, explicit public debt is much lower in China than in most other major economies while corporate debt is higher.

Figure 10. Components of Gross Debt



Source: IMF Global Debt Database.

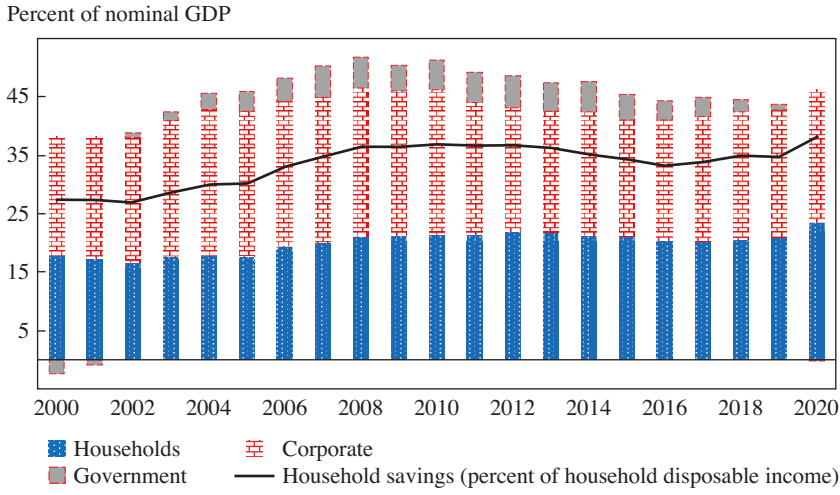
Corporate debt, which rose substantially in the aftermath of the global financial crisis, became a major source of concern and was seen as emblematic of China's investment-heavy growth model (figure 10). By 2016, nonfinancial corporate debt stood at 145 percent of GDP.[13] Recognizing these risks, the government initiated a deleveraging campaign that, by 2021, resulted in corporate debt falling back to 131 percent of GDP. Even though China has a high level of corporate debt, most of this is denominated in China's own currency and owned by domestic banks and investors, creating fewer fragilities than external debt (owed to foreign investors and denominated in foreign currencies such as the US dollar).

There are, however, specific sectors where the concentration of debt could be a problem. The real estate sector, in particular, is a broader source of economic vulnerability. Real estate investment has become a bulwark of the economy, helping to keep growth on an even keel even when other sectors flounder.[14] Local government officials are eager to sell land to developers,

13. International Monetary Fund, "Global Debt Database," https://www.imf.org/external/datamapper/datasets/GDD.

14. Rogoff and Yang (2021) estimate that the direct and indirect domestic value added of the real estate sector (both residential and commercial) is about 24 percent. Adding in the imported content of China's real estate sector would push that estimate up by about 3 percentage points. Rogoff and Yang (2022) and Huang (2023) highlight various other risks related to this sector and their systemic implications. Chen and others (2017) discuss how the real estate boom has exacerbated capital misallocation.

Figure 11. Components of Gross National Saving



Sources: CEIC and Flow of Funds, Ministry of Finance, China.
Note: Corporate saving calculated as the sum of gross savings of financial and nonfinancial institutions.

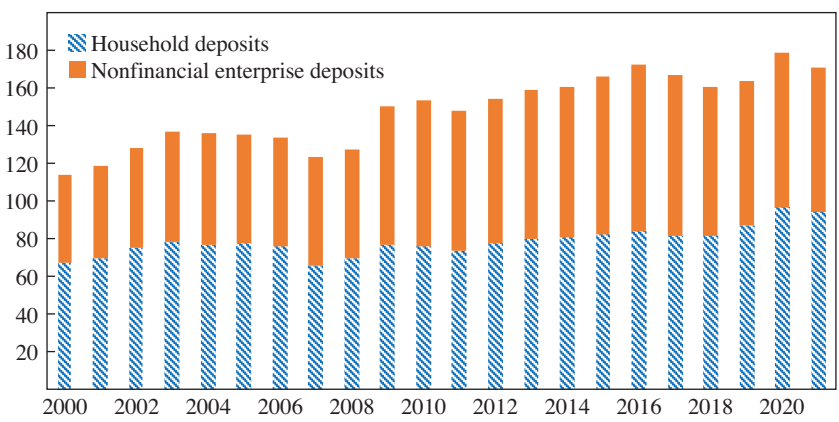
boosting government revenues and enabling a range of expenditures, including pet infrastructure projects. A fall in real estate prices or other factors that restrain real estate activity could thus have broad implications for growth through knock-on effects on other sectors, local government finances, and even household wealth.

Indeed, household exposure to the real estate sector, which has risen partly on account of government policies, has created additional vulnerabilities that could affect not just economic but also social stability. Easier access to residential mortgages, which the government encouraged, boosted housing demand and played a significant role in the surge in household debt, from about 30 percent of GDP a decade ago to slightly more than 60 percent of annual GDP (figure 10). Property has also become a mainstay of Chinese household wealth, accounting for nearly 60 percent of household wealth in 2019.[15] Thus, households are exposed in multiple ways to house price fluctuations.

While corporate and household debt levels are high, however, it is worth noting that saving rates are high as well (figure 11). Take, for instance, household saving which has averaged 25 percent of GDP in the last few

15. This estimate is based on a survey conducted by the People's Bank of China that is summarized in a news story, Xinhua (2020).

Figure 12. Banking Deposits: Household and Corporate



Sources: CEIC; Flow of Funds; and National Balance Sheet, China.

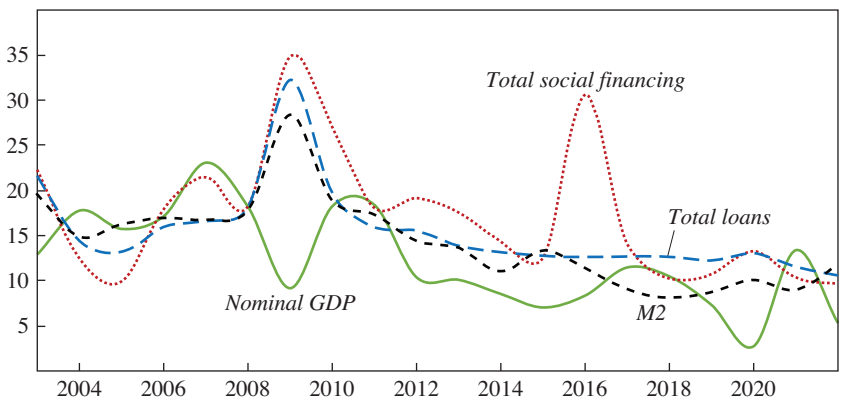
years. This is a flow variable as opposed to debt, which is a stock variable. A stock measure of household savings such as household deposits in the banking system, which amounted to 94 percent of GDP in 2021 (figure 12), portrays a more benign situation. In particular, overall household balance sheets do not appear at risk. In fact, the high level of household saving, which now amounts to 38 percent of household disposable income (figure 11), provides a measure of protection for household balance sheets but also inhibits the objective of boosting consumption growth.

One of the takeaways from this discussion is that much of the debt accumulation in China is financed by domestic savings, making it not so much a source of financial risk as of major inefficiencies and waste because of a broken system of allocating capital.[16] The state owns many of the key creditors and debtors, so a financial shock is unlikely to set off a financial crisis or a collapse in growth. There are still potential concerns about social stability if the government were to account for the true extent of losses in the banking system and use tax revenues to finance a recapitalization, which would be a large and complex undertaking.

Of course, it is not just aggregate levels of debt versus assets but how those are distributed through the economy that matter. A number of property

16. Song and Xiong (2018) reach a similar conclusion.

Figure 13. Growth in Nominal GDP and Monetary/Credit Aggregates

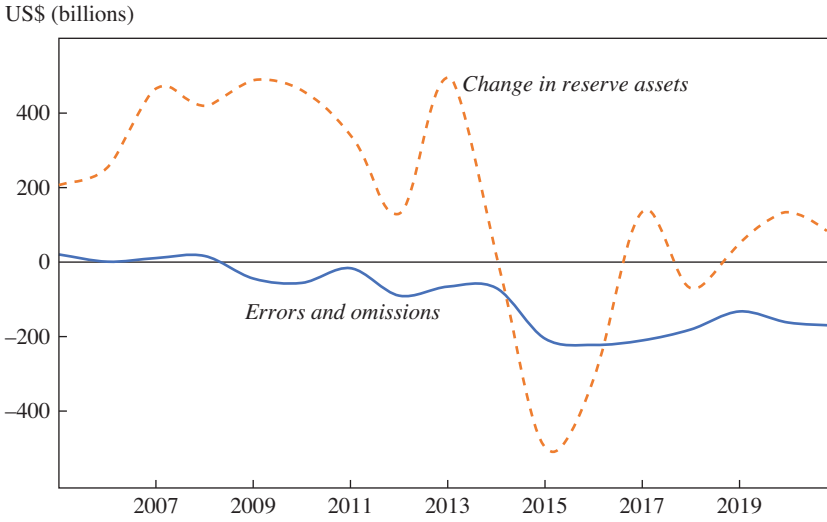


Source: CEIC.
Note: "Total loans" is the sum of total loans denominated in local currency and total loans denominated in foreign currencies.

developers, such as Evergrande Group, have run into financial trouble in recent years. Many property developers who are similarly exposed, with high debt levels and with vulnerable balance sheets if house prices were to tumble further, and a number of financial institutions that have lent to them are exposed to spillovers of problems from Evergrande. But a systemic meltdown is not in the cards. Most major Chinese banks are under state control and can provide infusions of cash to troubled corporations, even if that only pushes problems off into the future. Such stumbles are inevitable as China tries to give market forces freer rein, but the government has enough control and resources to prevent broader financial crashes.

On a less sanguine note, it is worth noting that for a number of years various monetary and credit aggregates have grown faster than nominal GDP (figure 13). These aggregates include M2 and measures of credit growth (total renminbi loans and total social financing, a broad measure of credit and liquidity that includes corporate bonds, foreign currency loans, trust loans, bank acceptance bills, and nonfinancial corporations' equity financing). There was a dramatic discrepancy between growth in these aggregates and in nominal GDP in 2009–2010, a period when the government unleashed a massive gusher of credit to support the economy by financing an investment boom (Hsieh, Bai, and Song 2016). Some of these discrepancies have persisted over time, including in 2020 and 2022. Coming on

Figure 14. Net Errors and Omissions and Changes in Reserves



Sources: CEIC and IMF.
Note: Changes in reserves are calculated from end-of-year stocks of foreign exchange reserves.

top of already high levels of leverage, this portends rising levels of debt that could pose problems in the future.[17]

II.B. External Vulnerabilities

Many emerging market economies have run into distress on account of high levels of external debt, particularly foreign currency debt than can cause balance sheet problems when a country's economy and exchange rate deteriorate simultaneously. China's overall external debt is estimated to be about 14 percent of GDP, far less than most other emerging market economies, and less than half of this is denominated in foreign currencies.[18]

Still, economic and political uncertainty have created concerns about capital flight. For a number of years now, the net errors and omissions item in the balance of payments has been substantially negative (figure 14).

17. According to official data from the China Banking and Insurance Regulatory Commission, the ratio of nonperforming loans to total loans (NPL ratio) in the banking system at the end of 2022 was 1.7 percent. Of the different categories of banks, the average NPL ratio was 1.3 percent for large commercial banks, 3.2 percent for rural commercial banks, and under 2 percent for all other categories. Whether these numbers are an accurate reflection of problem loans in the banking system and how large the effects of loan evergreening are remain open to question.

18. CEIC Data, https://www.ceicdata.com/en.

This is suggestive of a great deal of capital flowing out through unofficial channels, presumably to obviate controls on capital outflows.[19] These outflows seem to have picked up during 2015–2016, when the renminbi came under sharp depreciation pressures and the government also initiated an anti-corruption campaign, and have mostly stayed in the $150–200 billion range since then.

How large could these outflows be if, for instance, a banking panic were to lead to a flight of deposits out of the banking system and out of the country? In 2021, total deposits in the banking system amounted to about 170 percent of GDP (see figure 12). With foreign exchange reserves amounting to about 17 percent of GDP, flight of even 10 percent of total deposits would deplete reserves.[20] This is, of course, an unlikely scenario given that much of the banking system is state-owned and the government would probably back all deposits in the event of a financial panic. Moreover, capital control levers can be quite potent when the government directly controls much of the banking system and, therefore, the main conduits for large capital outflows.

Financial or currency meltdown scenarios are, almost by definition, the result of unexpected sequences of events. While these cannot be ruled out altogether, it is not easy to lay out a plausible scenario for such a meltdown in China.

III. Policy Frameworks and Contradictions

China has used a variety of organizing frameworks for its policies. These frameworks can often sound like slogans rather than policy statements. In fact, they serve as guideposts for a broad range of reforms and even serve to galvanize public support. While there have been reforms in recent years, many of these have been related to the financial sector and capital markets, with far less progress on supply-side and institutional reforms. This lack of balance creates its own risks.

19. The balance of payments identity is as follows: current account plus capital account plus net errors and omissions minus change in international reserves equals zero. A negative net errors and omissions implies that the sum of the current and capital accounts is positive (assuming no change in international reserves). During the early 2000s, when the renminbi faced substantial appreciation pressures and the government was trying to restrict capital *inflows*, net errors and omissions were persistently positive and large.

20. Obstfeld, Shambaugh, and Taylor (2010) argue that emerging market economies take M2 into consideration in determining their reserve holdings. The IMF now incorporates similar measures into its assessments of reserve adequacy.

III.A. Frameworks for Organizing Reforms and Reducing Resistance

Major economic reforms are often undertaken under the shadow of a crisis. Chinese leaders have undertaken some reforms even when short-term growth seems secure and despite potential risks and dislocations. There are multiple elements to any major reform effort—the framework, strategy, tactics, and implementation. At least on the first three, China has created an effective template.

To begin with, let us review examples of policy frameworks deployed by the government. Consider, for instance, how China rallied domestic support around the objective of making the renminbi an international currency, which in turn necessitated relaxing restrictions on cross-border capital flows.[21] Given its underdeveloped financial markets, capital account liberalization is premature for China. But this objective highlighted areas where reforms were in China's own interest—more efficient financial markets, better regulation of those markets, and a more flexible exchange rate. The framework proved important not only for catalyzing the individual reforms but in showing how they fit together and reinforce each other.

Second, a narrative that highlights the broad benefits of reforms can help blunt opposition. A major stumbling block to reforms in emerging market economies is the view, often a legitimate one, that such reforms largely benefit the economic and political elite. China has been effective at creating narratives that build broad support and provide a framework for communicating the logic and desirability of individual reforms. In early 2013, the Chinese government announced the common prosperity plan to reduce inequality, a dubious policy goal if it were to emphasize redistributive policies. In fact, the specific steps outlined in the plan were exactly the reforms China needed—financial market liberalization, reform of state-owned enterprises, and freer labor mobility. All worthy reforms in and of themselves but the narrative helped emphasize that the benefits were designed and intended to be widespread rather than accruing to the select few.

A clear timeline helps build credibility and reduce opposition to reforms. Some years ago, the government announced that within a two-year period it would eliminate the ceiling on interest rates paid on bank deposits. The ceiling had stifled bank competition, resulting in households getting minuscule inflation-adjusted returns on their deposits for much of the last decade.

21. See Prasad (2017), Eichengreen and Xia (2019), and Miao and Deng (2019). For a discussion of China's approach to capital account liberalization, see Prasad and Rajan (2008) and Clayton and others (2023). For a discussion of the indirect (collateral) benefits of capital account liberalization, see Kose and others (2009).

This sort of explicit commitment to a major policy change not only signaled seriousness but also blunted the opposition from banks, which had time to adjust to the change rather than having it foisted on them at short notice.

The implementation of such reforms is a fraught matter, however. China's approach often seems plodding and hypercautious, but it might have some merits in an economy beset with multiple inefficiencies. For instance, despite the strong rationale for the move, removing the ceiling on deposit rates had risks. Weaker banks could offer higher interest rates to compete for deposits, make riskier loans, and set themselves up for failure. While it put in place an explicit deposit insurance system to reduce these risks, the government chipped away at the deposit rate ceiling by allowing the proliferation of other saving products with higher returns. This approach had its own risks as "wealth management products" and off-balance-sheet lending proliferated. Regulators had to aggressively rein in these sorts of activities; nevertheless, these actions by the government eventually helped catalyze interest rate reforms.

Similarly, reform of state-owned enterprises has focused on subjecting them to harder budget constraints and market discipline rather than privatization. Keeping these enterprises in state hands lessens the imperative for reforms but also avoids the risks of large-scale privatization, such as the looting of assets. Small and indirect reform steps—even if inefficient overall relative to a big bang approach—elicit less opposition, pose fewer risks from the reforms themselves, and make course corrections easier.

The government has persisted with this approach. Recognizing the need to improve productivity and shift away from low-skill manufacturing, the government recently articulated the "dual circulation" policy, which implies continued engagement with global trade and finance but seeks a greater reliance on domestic demand as well as technological self-sufficiency and homegrown innovation. This policy, as well as the Made in China 2025 policy, have run into difficulties as China still needs foreign technology to upgrade its industry but US-China tensions have limited the availability of the necessary products and technological expertise.[22] Rising economic and geopolitical rifts with the West, particularly due to Beijing's stance on the Russian invasion of Ukraine and also because of the increasingly

22. Bergsten (2022) and Roach (2022) discuss the US-China economic rivalry and the possible fallout on both countries. For more discussion of the Made in China 2025 policy, as well as its impact and limitations, see Wei, Xie, and Zhang (2017) and Branstetter and Li (2022).

hostile environment toward foreign businesses operating in China, could more broadly limit China's access to foreign technology (and markets for its exports). In addition, there are limits to direct government intervention, not just in the allocation of resources but in an environment where reforms in complementary areas do not proceed in parallel.[23]

III.B. Unbalanced Reforms

Thanks to the efforts of some reform-minded officials, Xi Jinping's government seems to have understood the need for financial sector reforms and liberalization in order to promote better resource allocation. Fixing the financial system is not just about managing risks and avoiding financial disaster but also about allocating capital to the more productive, dynamic, and employment-generating parts of the economy. What will it take?

Since China's financial system is still bank dominated, fixing the banks is a key priority. The reported nonperforming loan ratios of major Chinese banks are low, but that is largely a reflection of that loan portfolio being concentrated in the state enterprise sector. Evergreening those loans, many of which may never be repaid in full, is in the interest of banks in order to keep those enterprises afloat and the loans from having to be recognized as nonperforming. Banks also need the right incentive structure to make loan decisions based on commercial considerations. Bankers are rational in responding to distorted incentives—loans to large state enterprises are safer even if those enterprises are insolvent because of implicit government guarantees. Solving these two problems requires recognizing and removing bad loans from banks' balance sheets, as well as reform of the state enterprises themselves, including weaning them off dependence on bank credit.

The government has also tried to improve market mechanisms to fortify the banking system by putting in place an explicit deposit insurance system, in the hope that markets would better discipline banks based on the riskiness of their balance sheets. But there appears to be a pervasive belief (garnered from observed yields and risk spreads) that the state still provides implicit full insurance to the entire banking system and perhaps even for other financial products. The government will need to demonstrate its seriousness about not protecting weak banks or unregulated saving products. (Much the same can be said for advanced economy governments and central banks!)

23. Branstetter, Li, and Ren (2022) find that Chinese industrial policies have not had salutary effects in promoting productivity.

In recent years, the government has had to deal with episodes of housing market and stock market volatility (Chong 2023).[24] The government has often found itself caught in a schizophrenic effort to strike a balance between maintaining confidence in the market and allowing the market to discipline itself, which has the perverse effect of heightening market turbulence. This on-off approach to intervention has sometimes injected a strong dose of uncertainty on top of already fragile investor sentiment and added to market volatility.

Another lesson to be taken from these episodes of volatility is not that market-oriented reforms can backfire but that they can add to volatility and generate more risks if they are not accompanied by broader reforms. China needs a better institutional framework—including more transparency in its policymaking process, better corporate governance and accounting standards, and more operational independence for the central bank and regulatory authorities—to supplement its financial and other market-oriented reforms. Progress in these areas has been lacking.

The government has rightly encouraged the development of stock and corporate bond markets, both as avenues for firms to raise financing and for savers to get higher returns and diversify their portfolios. But it has done little to improve corporate governance of Chinese companies or their accounting and auditing standards. The resulting opacity has contributed to large fluctuations in stock and bond markets, as investors have little information about the companies they are investing in, leading them to follow and exacerbate market swings. Although China has dropped most restrictions on capital inflows into its bond and stock markets, foreign investors remain wary for the reasons above and also because capital controls have been used as a tool to relieve large transitory pressures on the exchange rate (Prasad 2016).

This highlights the importance of getting regulatory frameworks and other institutional aspects such as public and corporate transparency right while freeing up markets. Otherwise, the result is likely to be more volatility and few of the benefits anticipated from such reforms.

When the government talks about state enterprise reform, it means subjecting them to market forces while maintaining state control. When it allows market forces a greater role in determining the exchange rate, it still wants to maintain stability and control. Reconciling these two contradictory impulses—more freedom for markets but with a heavy hand of

24. FRED Economic Data, "CBOE China ETF Volatility Index," https://fred.stlouisfed.org/series/VXFXICLS.

government intervention to maintain "stability and order"—poses difficult challenges.

Implementing even well-intentioned reforms in a second-best world with rampant inefficiencies involves transitional risks that might manifest in financial and economic volatility. The Chinese government has so far had enough resources and policy space to cope with some of those transitional risks, but it is a legitimate concern that its actions and attempts to clamp down on markets at difficult times might exacerbate problems, with long-lasting consequences.[25]

IV. Conclusion

China has found a way to get results—generating sustained growth over a long period, improving the living standards of its people, avoiding a financial crisis, and pulling its economy through a number of perilous periods for the world economy. It has done all of this without a well-functioning financial system, a strong institutional framework, a market-oriented economy, or a democratic and open system of government. There is certainly cause for humility for anyone attempting to explain the China phenomenon based on the historical record and experiences of other countries.

China's growth model and approach to reforms have not hewed to conventional norms and arguably tensions are building up in the system, with a possibly explosive meltdown at some point. But so far the government has proved adept at navigating around such perils. There have undoubtedly been mishaps, often with significant consequences, but the government has left itself room for maneuver. And there have been many resources wasted over time, with a big bill left to pay at some point in the future.

If the government's goal is to sustain growth, it needs to find ways to improve the allocation of resources within the economy and enhance productivity growth. This will require a better financial system. Indeed, while there are legitimate concerns about China's high rates of investment in

25. One interesting example is the People's Bank of China's attempt, in August 2015, to allow for a wider trading band for the renminbi's exchange rate relative to the dollar. This was intended as a move toward greater exchange rate flexibility. However, the central bank combined this policy change with a 1.9 percent step devaluation of the renminbi relative to the dollar. In the absence of clear communication from the central bank, this move was interpreted by markets as a signal of more devaluation to come in order to boost exports at a time when the economy was weakening. This set off worldwide turmoil in currency markets that lasted for over a year. It cost the Chinese central bank about $1 trillion in reserves to limit the renminbi's further depreciation and prevent a capital outflow/currency depreciation spiral.

physical capital, the capital-labor ratio is much lower than in advanced economies such as the United States, and China still has vast needs for infrastructure in its interior provinces. The challenge is the efficient intermediation of domestic savings into domestic investment, so capital is allocated to its most productive uses. China would benefit from a financial system that does a better job of allocating resources to more productive uses and to dynamic parts of the economy, especially the services sector and small and medium enterprises. This requires fixing the banking system, improving depth and liquidity in bond markets, and tightening regulation to mitigate institution-specific and system-wide risks. Such reforms, in tandem with institutional and supply-side reforms, will help reduce unproductive investment, improve employment and household income growth, and promote more regionally balanced development.

The underpinnings of China's growth seem fragile from historical and analytical perspectives. Things that must end do often end suddenly and in unpredictable ways. Yet, if the government plays its cards right, one could equally well envision a more benign future for the Chinese economy—with growth that is more moderate by its own standards, but that is more sustainable from economic, social, and environmental perspectives.

ACKNOWLEDGMENTS I thank Bineet Mishra, Caroline Smiltneks, Yicheng Yang, and especially Sharan Banerjee for excellent research assistance. Maurice Obstfeld, Kenneth Rogoff, and other *BPEA* conference participants provided useful comments.

References

Bailliu, Jeannine, Mark Kruger, Argyn Toktamyssov, and Wheaton Welbourn. 2017. "How Fast Can China Grow? The Middle Kingdom's Prospects to 2030." *Pacific Economic Review* 24, no. 2: 373–99.

Benzell, Seth, Laurence J. Kotlikoff, Maria V. Kazakova, Guillermo Lagarda, Kristina Nesterova, Victor Yifan Ye, and Andrey Zubarev. 2022. "The Future of Global Economic Power." Working Paper 30556. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30556.

Bergsten, C. Fred. 2022. *The United States vs. China: The Quest for Global Economic Leadership*. Cambridge, Mass.: Polity Press.

Bosworth, Barry, and Susan M. Collins. 2008. "Accounting for Growth: Comparing China and India." *Journal of Economic Perspectives* 22, no. 1: 45–66.

Branstetter, Lee G., and Guangwei Li. 2022. "Does 'Made in China 2025' Work for China? Evidence from Chinese Listed Firms." Working Paper 30676. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30676.

Branstetter, Lee G., Guangwei Li, and Mengjia Ren. 2022. "Picking Winners? Government Subsidies and Firm Productivity in China." Working Paper 30699. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30699.

Chen, Ting, Laura Liu, Wei Xiong, and Li-An Zhou. 2017. "Real Estate Boom and Misallocation of Capital in China." Working Paper. https://economics.princeton.edu/working-papers/real-estate-boom-and-misallocation-of-capital-in-china/.

Chen, Wei, Xilu Chen, Chang-Tai Hsieh, and Zheng (Michael) Song. 2019. "A Forensic Examination of China's National Accounts." *Brookings Papers on Economic Activity*, Spring, 77–141.

Chong, Andrew. 2023. "China Real Estate: Still Very Relevant on Earth." J.P.Morgan Asset Management, May 30. https://am.jpmorgan.com/us/en/asset-management/adv/insights/portfolio-insights/fixed-income/fixed-income-perspectives/china-real-estate-still-very-relevant-on-earth/.

Clark, Hunter, Maxim Pinkovskiy, and Xavier Sala-i-Martin. 2020. "China's GDP Growth May Be Understated." *China Economic Review* 62:101243.

Clayton, Christopher, Amanda Dos Santos, Matteo Maggiori, and Jesse Schreger. 2023. "Internationalizing Like China." Discussion Paper 17781. London: Centre for Economic Policy Research. https://cepr.org/publications/dp17781.

Dollar, David, Yiping Huang, and Yang Yao, editors. 2020. *China 2049: Economic Challenges of a Rising Global Power*. Washington: Brookings Institution Press.

Eichengreen, Barry, Donghyun Park, and Kwanho Shin. 2012. "When Fast-Growing Economies Slow Down: International Evidence and Implications for China." *Asian Economic Papers* 11, no. 1: 42–87.

Eichengreen, Barry, and Guangtao Xia. 2019. "China and the SDR: Financial Liberalization through the Back Door." *Quarterly Journal of Finance* 9, no. 3: 1950007.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105, no. 10: 3150–82.

Fernald, John G., Eric Hsu, and Mark M. Spiegel. 2021. "Is China Fudging Its GDP Figures? Evidence from Trading Partner Data." *Journal of International Money and Finance* 110:102262.

Holz, Carsten A. 2014. "The Quality of China's GDP Statistics." *China Economic Review* 30: 309–38.

Hsieh, Chang-Tai, Chong-En Bai, and Zheng (Michael) Song. 2016. "The Long Shadow of a Fiscal Expansion." *Brookings Papers on Economic Activity*, Fall, 129–65.

Huang, Tianlei. 2023. "Understanding China's Housing Problems." Unpublished manuscript. Washington, DC: Peterson Institute for International Economics.

International Monetary Fund (IMF). 2022. *World Economic Outlook: Countering the Cost-of-Living Crisis*. Washington: IMF. https://www.imf.org/en/Publications/WEO/Issues/2022/10/11/world-economic-outlook-october-2022.

Jurzyk, Emilia M., and Cian Ruane. 2021. "Resource Misallocation among Listed Firms in China: The Evolving Role of State-Owned Enterprises." Working Paper 2021/75. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2021/03/12/Resource-Misallocation-Among-Listed-Firms-in-China-The-Evolving-Role-of-State-Owned-50167.Drop

Kharas, Homi, and Indermit S. Gill. 2020. "Growth Strategies to Avoid the Middle-Income Trap." Social Science Research Network, January 28. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3526261.

Kose, Ayhan, Eswar Prasad, Kenneth S. Rogoff, and Shang-Jin Wei. 2009. "Financial Globalization: A Reappraisal." *IMF Staff Papers* 56, no. 1: 8–62.

Lardy, Nicholas R. 2014. *Markets over Mao: The Rise of Private Business in China*. Washington: Peterson Institute for International Economics.

Lardy, Nicholas R. 2019. *The State Strikes Back: The End of Economic Reform in China?* Washington: Peterson Institute for International Economics.

Lardy, Nicholas R. 2020. "China Has Been Overstating the Role of Private Investment in Its Economy." Blog post, March 5, Peterson Institute for International Economics. https://www.piie.com/blogs/china-economic-watch/china-has-been-overstating-role-private-investment-its-economy.

Miao, Yanliang, and Tuo Deng. 2019. "China's Capital Account Liberalization: A Ruby Jubilee and Beyond." *China Economic Journal* 12, no. 3: 245–71.

Obstfeld, Maurice, Jay C. Shambaugh, and Alan M. Taylor. 2010. "Financial Stability, the Trilemma, and International Reserves." *American Economic Journal: Macroeconomics* 2, no. 2: 57–94.

Prasad, Eswar S. 2016. "Capital Controls Will Not Cure China's Ills." *Financial Times*, January 31. https://www.ft.com/content/c1bf7efe-c67b-11e5-808f-8231cd71622e.

Prasad, Eswar S. 2017. *Gaining Currency: The Rise of the Renminbi*. New York: Oxford University Press.

Prasad, Eswar S., and Raghuram G. Rajan. 2008. "A Pragmatic Approach to Capital Account Liberalization." *Journal of Economic Perspectives* 22, no. 3: 149–72.

Pritchett, Lant, and Lawrence Summers. 2014. "Asiaphoria Meets Regression to the Mean." In *Prospects for Asia and the Global Economy*, edited by Reuven Glick and Mark M. Spiegel. San Francisco: Federal Reserve Bank of San Francisco.

Roach, Stephen. 2022. *Accidental Conflict: America, China, and the Clash of False Narratives*. New Haven, Conn.: Yale University Press.

Rogoff, Kenneth S. 2022. "China's Diminishing Returns." *Project Syndicate*, October 31. https://www.project-syndicate.org/commentary/china-diminishing-returns-real-estate-housing-slowdown-by-kenneth-rogoff-2022-10.

Rogoff, Kenneth S., and Yuanchen Yang. 2021. "The Size of China's Real Estate Sector." https://scholar.harvard.edu/sites/scholar.harvard.edu/files/rogoff/files/the_size_of_chinas_real_estate_sector.pdf.

Rogoff, Kenneth S., and Yuanchen Yang. 2022. "A Tale of Tier 3 Cities." Working Paper 30519. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30519.

Rosen, Daniel H. 2021. "China's Economic Reckoning." *Foreign Affair*s, June 22. https://www.foreignaffairs.com/articles/china/2021-06-22/chinas-economic-reckoning.

Sasaki, Takatoshi, Tomoya Sakata, Yui Mukoyama, and Koichi Yoshino. 2021. "China's Long-Term Growth Potential: Can Productivity Convergence Be Sustained?" Working Paper 21-E-7. Tokyo: Bank of Japan. https://www.boj.or.jp/en/research/wps_rev/wps_2021/wp21e07.htm.

Setser, Brad W. 2018. "China's Currency Is Back in Play." Blog post, July 23, Council on Foreign Relations. https://www.cfr.org/blog/chinas-currency-back-play.

Song, Zheng (Michael), and Wei Xiong. 2018. "Risks in China's Financial System." *Annual Review of Financial Economics* 10:261–86.

Wei, Shang-Jin, Zhuan Xie, and Xiaobo Zhang. 2017. "From 'Made in China' to 'Innovated in China': Necessity, Prospect, and Challenges." *Journal of Economic Perspectives* 31, no. 1: 49–70.

Xinhua. 2020. "Average Net Worth of Chinese Urban Household Nears 3 Mln Yuan: Survey." XinhuaNet, April 26. http://www.xinhuanet.com/english/2020-04/26/c_139009659.htm.

Comment and Discussion

COMMENT BY

KENNETH ROGOFF It is a pleasure to discuss these three superb papers on India, Brazil, and China. India and China, of course, both have populations over 1.4 billion people.[1] Brazil is the most important economy in Latin America. All three papers deserve immense credit for managing to give a sweeping overview of current macroeconomic issues facing these economies in very condensed form.

Although these economies have been grouped together in a single panel, and they are all important socially, economically, and politically, China is vastly the most significant in the global economy. It is by any measure the world's second-largest economy, and by some measures first. Using end of 2021 World Bank data, China's 2021 GDP in current dollars at market exchange rates was $17.8 trillion.[2] India, with roughly the same size population of 1.4 billion people, had a GDP of only $3.2 trillion. Brazil's economy is only about half the size of India's, but then it has only approximately 15 percent as many people. India becomes significantly larger if one uses the Penn World Table (PWT) to compare "expenditure equivalent" per capita GDP; this is a way to try to compare incomes using a common set of prices for different goods and services. PWT 10.1 is the most recent version, which runs through 2017. It gives per capita GDP in China as roughly $13,700, in India $6,000, and Brazil $14,300; for 2017, the United States is $60,800.[3]

1. World Bank, "Population, Total," https://data.worldbank.org/indicator/SP.POP.TOTL.
2. World Bank, "GDP (Current US$)," https://data.worldbank.org/indicator/NY.GDP.MKTP.CD.
3. University of Groningen, "Groningen Growth and Development Centre," https://www.rug.nl/ggdc/productivity/pwt/?lang=en.

Brookings Papers on Economic Activity, Spring 2023: 271–288 © 2024 The Brookings Institution.

A BRIEF OVERVIEW Acharya offers a broad perspective on why India has failed to grow faster and what economic changes are needed to do better. He focuses especially on how Indian growth has been shackled by monopolies. A few conglomerates, particularly the top handful, have outsized footprints and influence. The concentration would not necessarily be crippling if, as in Korea, these firms had to compete with world-beaters from other countries. But the Indian political class's insistence on maintaining perhaps the world's largest tariff barrier regime coddles and protects highly inefficient firms. Heavy protectionism stunts growth and makes India a far less central player in the global economy than it should be. One hopes this will change over time, but so far political forces have resisted mightily, despite the obvious success of the IT and outsourcing sectors, not to mention pharmaceuticals. No single reform would be a magic bullet for development, and Acharya emphasizes several, especially improving education and expanding opportunities for women. Although India has managed decent growth in recent years—indeed, somewhat faster than China—it has been heavily concentrated in urban areas, with rural India still left far behind.

In contrast to India, which has produced some very wealthy people but remains a lower-middle-income nation, China's growth has been the envy of the world, albeit recent rethinking of the way economists compare growth across countries makes China's performance slightly less impressive than had previously been thought. Prasad's paper is replete with important nuances and empirical refinements, making it an invaluable resource. This reappraisal does not diminish the fact that China has continually managed to outperform the expectations of many Western economists who have long seen China as an accident waiting to happen. Given its seemingly weak institutions, inefficient financial sector, and the slow pace of reform, China's performance is remarkable indeed. Prasad acknowledges the arguments of skeptics but pushes back forcefully on the view that a financial crisis or a major slowdown in trend growth is likely in the foreseeable future; the Chinese authorities have simply proven too adept, again and again.

I learned a lot from the paper, but as I shall discuss below, I am considerably less sanguine about the future. Most immediately, China is likely running into sharply diminishing returns to investment in real estate and infrastructure, as documented in the city-by-city analysis of Rogoff and Yang (2022), which is worrisome indeed given the reliance on real estate construction and infrastructure investment for growth (Rogoff and Yang 2021). The overbuilding problem is particularly acute in the so-called tier 3 cities and below, which are smaller and poorer than the tier 1 and

tier 2 cities, yet nevertheless collectively account for 60 percent of China's GDP. In fact, in Prasad's commendably comprehensive analysis, he does note that although China's capital labor ratio is only 28 percent that of the United States, its capital-to-output ratio considerably *exceeds* that of the United States. There is also the matter of China's rapidly aging population and the overcentralization of economic control, which has become considerably more acute the past decade. Last but not least, China appears to have increasingly emphasized investment in defense and security over growth.

The paper by Carvalho and Nechio on Brazil is much narrower in scope than the other two papers but is not lacking in importance, rigor, and ambition. The paper exploits a remarkable data set on the inflation expectations of professional forecasters that the central bank in Brazil, Banco Central do Brasil (BCB), with tremendous foresight, started gathering in 2002, shortly after the country's adoption of inflation targeting. The data set allows the authors to construct a measure of "inflation unanchoring" with medium-term expectations (they particularly focus on eighteen months in line with the BCB's focus). They show that there is now a worrying unmooring of inflation expectations, which has taken place twice before in this century, in each case only being reversed with painful adjustments, and only when fiscal policy became supportive. They show that "reanchoring" typically takes a couple of years and at large cost. Although certainly Brazil is quite different than an advanced economy today, the authors argue that nevertheless the estimates are sobering, and of course it was not so long ago that advanced economies experienced enormous difficulties in bringing down inflation expectations.

FURTHER THOUGHTS Acharya gives an excellent overview of where India's economy stands today, not only providing cogent analysis of the problems but a useful quantitative perspective. A key theme is that India's high aggregate growth numbers in recent years mask how little progress has been made in rural India, where, as Acharya observes, well over half the population lives at barely subsistence levels. (Academic debate about exactly where the world poverty line lies obscures the almost incomprehensible chasm between the standard of living for low-income residents of Western nations and low-income citizens of rural India.) Acharya argues that without key reforms, things are not going to improve anytime soon. Among the reforms he emphasizes are lowering tariffs, reducing monopoly power and fiscal deficits, improving primary and secondary education, and raising the labor force participation of women. Although all of these are of fundamental importance, perhaps the most frustrating are India's tariff barriers, since of course they can be reduced with the stroke of pen. Tariffs

for most goods did come down sharply thanks to India's hugely successful 1991 IMF program, but with an average rate of over 18 percent in 2021, they are still among the highest in the world (WTO 2022; Alessandrini and others 2009). India's tariff wall effectively cuts it off from the global supply chains which have bolstered growth among many other Asian countries in recent decades. Supply chains work by assigning production to the most efficient supplier country at every stage, but they cannot work if every time an intermediate product crosses a national border, it is hit with a heavy tariff. India's tariff regime is perhaps built for another era, certainly not this one.

For years, defenders of the high-tariff status quo have pushed back, arguing that as modern manufacturing plants become ever more automated, the time to follow the Asian growth model as a vehicle to employ India's youthful population had long passed. While it is true that technology and robotics are increasingly making employment growth through manufacturing exports a less viable approach in some industries, countries such as Vietnam are showing there are still plenty of lower-tech industries where it can be done. Moreover, just now, with geopolitical tensions forcing many firms to adopt a China Plus One model—that is, to keep existing plants in China but diversify with plants in another country for future production growth—India would have a huge opportunity were it not isolated by tariffs and other forms of protectionism, for example, limits on foreign ownership and investment.

Longer term, for India to reach its potential, it must also educate its rural population and bring more women in the labor force; Acharya shows that on both counts, India does not stack up well against successful Asian economies. With the right policies, there is the potential to bring hundreds of millions of desperately poor Indian citizens into the world's middle class. True, the one percent in India would hugely benefit from changes that raise India's growth trajectory, leading many rich-country academics and media commentators to complain about neoliberalism and the Washington Consensus. My guess, however, is that the overwhelming majority of Indians would vote for freeing up India's economy in a heartbeat, if the costs, benefits, and risks were honestly and transparently presented to them. (It would be important to redistribute a share of the benefits of globalization, an area where the United States has fallen short.) Unfortunately, there are deep-seated reasons why India has not managed to achieve the growth seen in East Asia.

In his passionately argued 2023 book, *India Is Broken,* Ashoka Mody argues that since achieving independence in 1947 India's government, beginning with its first prime minister, Jawaharlal Nehru, has failed its

people. Nehru and later prime minister Indira Gandhi famously eschewed markets for a form of democratic socialism that worked to a degree at first but that by the mid-1980s had completely lost credibility. As has often been the case with overly powerful centralized governments, absent strong checks and balances, a system of cronyism and corruption first crept in and then became all-consuming. Powerful monopolies, protected by a massive wall of trade barriers, became ingrained. For the (relatively) rich Indians, these were not a flaw but a feature of the system, and they successfully lobbied to protect it, as Mody demonstrates. India failed to ride the manufacturing export wave that first Japan, then the Four Asian Tiger countries, and then China and Vietnam, have exploited.

As noted earlier, India did open up in the 1990s, prodded by a transformational IMF program (most Indian policymakers give the IMF program considerable credit even if some of the IMF's most vocal Western academic critics cannot bring themselves to do so). Indeed, the IMF program was hardly radical, but nevertheless it led to reforms that gradually propelled India forward for some time. Unfortunately, this initiative had burned out by the early 2000s. Today, India, with Narendra Modi as its current president, has found growth in some areas—particularly in digital technology, as Acharya documents—but the majority of the population has been left behind, living at a level of rural poverty most Americans have never seen and have no concept of.[4] Whatever one thinks of the American economy having not produced good middle-class jobs, the problem is an order of magnitude worse in India. With its incredibly young population (more than half under thirty), India ought to be taking over as the growth engine of the world (Pradhan and Chatterjee 2023). Instead, even college graduates, if they fail to land cushy lifetime government jobs or ambitious jobs in the digital sector, are either unemployed or massively underemployed. India has tried to emulate China to some extent by investing heavily in infrastructure, and there have been massive improvements, but the sector is perhaps less efficient than China's; moreover, it is difficult to reap the rewards as long as the tariff wall stands.

To be clear, Mody's (2023) prescription of having Indian politicians be more honest and less self-serving seems utopian. But his diagnosis of the core problem—India's politicians and political system—is spot on, and how to fix this cuts to the heart of development. There is a huge debate in development about whether the problem is institutions—as Acemoglu

4. Oxfam International, "India: Extreme Inequality in Numbers," https://www.oxfam.org/en/india-extreme-inequality-numbers.

and Robinson (2012), building on Landes (1998), argue—or culture, which Alesina and Guiliano (2015) emphasize. India has suffered from having prime ministers who took on too much power and did not use it wisely, with perhaps the worst offender being Indira Gandhi in the 1970s. Yet this argument alone does not provide a sufficient explanation; other countries have had strong autocratic leaders who nevertheless managed to open up their economies to benefit from trade integration.

The postwar experience speaks very clearly to the point that any countries that have tried an insular approach and eschewed strong export orientation have failed to achieve development escape velocity, unless one wants to point to the former Soviet Union.

Acharya notes that India's consolidated government deficit has consistently exceeded 10 percent in recent years, reaching as high as 14 percent. Of course, some of this is simply the result of the pandemic, but it has been a long-standing problem in India for decades now. Historically, the government stuffed the debt into the state-controlled banks, which allowed the deficits to be financed at relatively low interest rates but at the cost of crowding out private investment. Acharya notes that today the capital markets are considerably more liberalized so the risks of financial repression are less. But this is far truer of equity than debt markets precisely because the bankruptcy laws are so weak and in need of repair. More complete financial liberalization, even if just domestically, would unquestionably put the debt—which exceeds 83 percent of GDP according to the October 2022 IMF Fiscal Monitor—under far greater pressure. Again, the weakness of the political system has made the debt problem chronically difficult to fix, and it has only been contained by constraining India's capital markets and capital market openness. The one bright spot is that reserves now exceed $550 billion, making it relatively easy for the central bank to stabilize exchange rates.

One important area Acharya does not touch on is India's energy transition, which is likely to be a huge point of contention between India and the West for decades to come and to imply massive costs that need to be addressed. India is certainly taking great efforts to use wind and especially solar and is steadily building nuclear power plants, which many advanced economies have perhaps unwisely sidelined (Goldstein and Qvist 2019). India is also heavily reliant on coal, however, which is cheap and abundant, and the government is reluctant to abandon coal for reasons of energy security. It seems to me that the United States and Europe have yet to reconcile themselves to the fact that their own green energy transition efforts, which are especially important to show leadership and potentially to find

new technologies, will not be nearly enough to prevent significant rises in global temperatures without inducing developing economies to also embrace the green energy transition. In many cases, and particularly in the case of India, developing economies will need strong inducement, not just financing. In Rogoff (forthcoming), I expand on a long-standing proposal I have had for establishing a world carbon bank, precisely to systemically help developing economies with the green energy transition, providing a framework for both economic incentives and technical assistance.

Prasad's analysis of China asks whether, after the pandemic, China will be able to resume anything like the kind of growth rates the country has had in the past. The paper offers a thoughtful and balanced overview of the Chinese economy. Prasad is absolutely right to emphasize how Western economists need to have some humility in assessing China's growth prospects, given how it has managed to persist with high growth despite low levels of financial and institutional development, high levels of debt, and a state-dominated economy.

However, my own take is more pessimistic. I completely agree that the problem China faces is not necessarily imminent financial catastrophe; Prasad dismisses that given China's extremely high household savings rate (over 25 percent), relatively low external debt, and especially the ability of the authorities to quickly intervene and allocate losses in the event of any financial crisis. Admittedly, this relatively sanguine view of China's financial vulnerabilities has yet to be stress-tested. Despite having a central government that can adjudicate losses far more expeditiously than say, the United States' court system, a financial shock would cause huge battles over the size and allocation of losses. And even absent bank runs and bank collapses, it would be exceedingly difficult for China to avoid the credit collapse that would inevitably accompany a sharp weakening of growth, with central government stretched thin to bail out both financial firms sitting on bad debts and heavily indebted local governments.

Setting aside crisis, China still faces sharply diminishing returns on its go-to growth sector, real estate. Including direct and indirect inputs, real estate accounts for roughly a quarter of total aggregate demand in China, and that figure would be even higher still if infrastructure spending is included (Rogoff and Yang 2022). The core problem is not just the overall size of the sector but the fact that cumulative construction and building is so huge after decades of building at breakneck speed.

In Rogoff and Yang (2022), we look at detailed city region–by–city region construction and real estate data for 284 Chinese cities, ranging from the four marquee tier 1 cities (Beijing, Shanghai, Shenzhen, and

Guangzhou) to the thirty-five tier 2 cities to 250 smaller and generally less affluent tier 3 cities, over the period 2000 to 2022, including both residential and commercial real estate. We reach two key conclusions. First, the size of the real estate sector in China is enormous; as a percentage of GDP, it is on par with Ireland and Spain at the peak of their booms, and far higher than that of the United States before the crash that began in 2007. The second conclusion comes from constructing a novel annual measure of the total cumulative size of the housing stock, making use of national census data (every ten years) and annual data available at the city level. Incredibly, China appears to now have a measure of square meters per capita of housing on par with countries such as France, Germany, and the United Kingdom, which have per capita income (measured at current dollar exchange rates) more than three times that of China's. The broader point is that after years of massive building of real estate and infrastructure, China is almost surely running into diminishing returns. The problems are particularly acute in tier 3 cities, which account for 60 percent of China's GDP and roughly 50 percent of the market value of its real estate. In effect, China has poured money into construction in tier 3 cities on the theory that it would help incentivize and rationalize a greater spreading out of the population, which can be observed in most countries. But simply put, the good jobs never came, and the tier 3 cities are now experiencing falling housing prices and declining populations. The fact that this key driver of China's growth strategy is running into diminishing returns is reminiscent of the obstacles faced by the investment-led growth strategies of Russia, Japan, and East Asia.

And financial risks cannot be dismissed entirely. It is true that the state-owned banks are basically backed by the central government, but city-level commercial banks and rural commercial banks have been rising rapidly in importance, particularly after 2005. They are more aggressive than bigger state-owned banks and have arguably become a significant risk factor in the banking system. Moreover, China is a bank-dominated financial system, where financing through bank loans is much larger than financing through capital markets. The shift to direct financing through bonds, as China aspires to do to enhance efficiency, will not be easy.

There are other problems as well. China faces a demographic crater: its population is projected to fall from over 1.4 trillion people today to a trillion or less by 2100.[5] As Prasad notes, the labor force is set to fall

5. United Nations, "Population," https://www.un.org/en/global-issues/population.

by 1 percent per year starting in 2030. China still has much to gain from importing Western technology, again as noted by Prasad, but surely there are diminishing returns here too compared to the early decades of opening up. Perhaps the most problematic near-term problem though is the huge centralization of power and decision-making that has taken place since 2013 under Chinese president Xi Jinping. Part of the reason China has been able so successfully to navigate problems that have derailed other fast-growing emerging markets is the country's ability to have very effective technocrats, making effective use of competition, incentives, and continuing education. It is not clear whether this tradition of excellence will survive the increasing centralization of power. If the desire to emphasize loyalty over competence weakens China's economic management, obviously the country's ability to consistently outperform will weaken.

Finally, how much has China actually outperformed? Prasad presents some interesting new numbers that have come out of attempts to develop more refined cross-country comparison of growth. He notes that "in the latest version of the PWT (10.01), China's growth looks a little less remarkable. . . . Average annual per capita GDP growth over 1977–2010 . . . is 8.75 based on PWT 8.0 data and roughly 6 percent based on PWT 9.0 and 10.01 (which has data through 2019). China's per capita GDP growth rate was 3.9 percent from 2011 through 2019 per PWT 10.01 and 6 percent from 2011 through 2022 in the official data." This is actually a rather stunning downgrade of China's performance, and it is an important contribution by Prasad to highlight it. It's beyond the scope of this comment to discuss the details, but fundamentally it is not trivial to try to make purchasing power parity comparisons of output across countries that have vastly different production and consumption patterns.

We now turn to Brazil. The paper by Carvalho and Nechio focuses narrowly on one particular problem that Brazil faces, anti-inflation credibility. Moreover, Brazil's inflation credibility problem is emblematic of a host of broader macroeconomic problems that it faces. The paper begins with the real exchange rate stabilization problem in the early 1990s, but to put the country's credibility problems in context, it is important to understand that Brazil had two periods of hyperinflation before that (Reinhart and Rogoff 2009). After the collapse of the real plan, Brazil instituted an extremely ambitious inflation targeting regime in 1999 under central bank governor Arminio Fraga, who had been airlifted into the government from his position as a hedge fund manager in New York. Along with the inflation targeting regime instituted by Mexico's Guillermo Ortiz Martínez, this was the first time inflation targeting—really a mechanism for asserting central bank

independence—had been tried seriously in an emerging market. The enduring success of the regime is remarkable; the fact that we can now consider it a problem when inflation expectations deviate by even 2 or 3 percentage points from target is quite an achievement in a country that has known 2,000 percent inflation (Reinhart and Rogoff 2009, 186). Of course, inflation targeting would be meaningless without central bank independence to underpin it (Rogoff 1985), and here the Banco Central do Brasil has been remarkably effective in asserting itself, despite incredible pressures at times.

The authors mark three instances where despite the inflation targeting regime, credibility became de-anchored in the way they attempt to measure. All three times were associated with political transition: 2002 with the first election of Lula, 2015 with the second term of Rousseff, and now again starting in 2021 with the return of Lula. They emphasize how, in all three cases, political transition sparked fears of a general loosening of fiscal policy combined with an undermining of central bank independence. In the first two cases, reanchoring inflation expectations (by their measure) took more than two years and, they argue, came at great cost to output.

The current de-anchoring began under President Bolsonaro, who came into office emphasizing fiscal discipline but who first, in response to COVID-19 and to lagging in the polls against Lula, expanded deficits significantly. It is remarkable how successfully Brazil was able to do this without exploding interest rates, and certainly this would not have been possible without the public believing that the Banco Central do Brasil enjoyed significant independence. Things became worse when Lula won the election and even more so when he came to office, with policies that involved even further expanding deficits. Unlike in advanced economies, Latin American economies were relatively quick to start raising interest rates as inflation unfolded, with Brazil's central bank raising its benchmark Selic rate to over 13 percent.[6] Curiously, the authors find that despite the explosion of inflation Brazil nevertheless experienced, and the obviously expansive fiscal policy, inflation expectations did not become unanchored right away, and really became so only as Lula came to power. Even more curiously, inflation expectations started becoming unanchored after inflation had peaked and seemed to be getting under control. It is an interesting example of how slowly inflation expectations adjust, and if there is a lesson for advanced economies, it is that even if inflation expectations had only

6. Banco Central do Brasil, "Selic Interest Rate," https://www.bcb.gov.br/en/monetarypolicy/selicrate.

moved slightly so far, they may move farther even as inflation falls below its peaks, making it difficult to go the last mile in reducing inflation as central banks hope to do.

A complication in interpreting the Brazilian experience is that its real interest rates have tended to be among the highest of all major emerging markets (along with Turkey) frequently in the past (Seguro-Ubiergo 2012). Brazil's high real interest rates may have to do partly with lingering default concerns but likely also capture an inflation risk premium that may not be well captured in the mean forecasts of the market. The risk premium to compensate for inflation risk, above and beyond what is captured in expected inflation, has likely fallen steadily in advanced economies over the past couple of decades. Another lesson from Brazil is that even if inflation expectations remain reasonably anchored, it is quite possible—indeed perhaps likely—for the term premium in longer-term bonds to increase, even in advanced economies.

In sum, this is an excellent set of papers, giving concise and interesting insights to three extremely important economies.

REFERENCES FOR THE ROGOFF COMMENT

Acemoglu, Daron, and James Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty.* New York: Crown.

Alesina, Alberto, and Paola Guiliano. 2015. "Culture and Institutions." *Journal of Economic Literature* 53, no. 4: 898–944.

Alessandrini, Michele, Bassam Fattouh, Benno Ferrarini, and Pasquale Scaramozzino. 2009. "Tariff Liberalization and Trade Specialization in India." Working Paper 177. Manila: Asian Development Bank. https://www.adb.org/sites/default/files/publication/28258/economics-wp177.pdf.

Goldstein, Joshua S., and Staffan A. Qvist. 2019. *A Bright Future: How Some Countries Have Solved Climate Change and the Rest Can Follow.* New York: Public Affairs.

Landes, David S. 1998. *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor.* New York: W. W. Norton.

Mody, Ashoka. 2023. *India Is Broken: A People Betrayed, Independence to Today*. Palo Alto: Stanford University Press.

Pradhan, Bibhudatta, and Adrija Chatterjee. 2023. "What India Becoming the World's Most-Populous Nation Means." *Washington Post*, April 19. https://www.washingtonpost.com/business/2023/04/19/explainer-what-india-passing-china-to-become-world-s-most-populous-nation-means/a434f924-de90-11ed-a78e-9a7c2418b00c_story.html.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly.* Princeton, N.J.: Princeton University Press.

Rogoff, Kenneth S. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics* 100, no. 4: 1169–89.

Rogoff, Kenneth S. Forthcoming. "The Case for a World Carbon Bank." *Journal of Policy Modeling* 45.

Rogoff, Kenneth S., and Yuanchen Yang. 2021. "Has China's Housing Production Peaked?" *China and World Economy* 29, no. 1: 1–31.

Rogoff, Kenneth S., and Yuanchen Yang. 2022. "A Tale of Tier 3 Cities."Working Paper 30519. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30519.

Seguro-Ubiergo, Alex. 2012. "The Puzzle of Brazil's High Interest Rates." Working Paper 2012/062. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2016/12/31/The-Puzzle-of-Brazil-s-High-Interest-Rates-25751.

World Trade Organization (WTO). 2022. *World Tariff Profiles 2022*. Geneva: WTO, ITC, and UNCTAD. https://www.wto.org/english/res_e/booksp_e/world_tariff_profiles22_e.pdf.

GENERAL DISCUSSION Olivier Blanchard made a comment on the significance of whether the Big-5 industrial conglomerates mentioned in Viral Acharya's paper have become so big that it would be politically extremely difficult to decrease tariffs.

Acharya highlighted this as a key point, suggesting that culture and institutions are linked and that autocratic regimes weaken institutions by attempting to show their ineffectiveness, instigating a shift from reliance on institutions to reliance on the leader. He posited that the focus of the leader on being able to get credit for the progress of the country presents the root of the political economy issue. Acharya referred to Kenneth Rogoff's mention of *India Is Broken* by Ashoka Mody, concluding that this has been the case in India.[1] He expanded on the point, describing how India was a reasonably private economy under the guise of a socialist economy until Indira Gandhi nationalized a large portion of India. Acharya suggested that she did so in an effort to get reelected by earning credit for the progress made in the country that was being attributed to private companies. Acharya also suggested that since this nationalization approach has failed, India has turned to selecting four or five national champions in the economy as its industrial policy.

1. Ashoka Mody, *India Is Broken: A People Betrayed, Independence to Today* (Palo Alto: Stanford University Press, 2023).

In response to Rogoff's discussant remarks, Acharya noted that India should be prepared for the potential bankruptcy of a large conglomerate and his paper aimed at proposing adequate policy measures. He pointed out that foreign borrowing costs of some of the large conglomerates in India have risen significantly and as a result they are deleveraging sharply, which he suggests is a sign that there might be a problem of a credit risk.

Kristin Forbes referenced Pinelopi Goldberg's paper on whether the world is deglobalizing, also presented at the Spring *BPEA* Conference. Forbes posed the question to each of the panelists of whether the risks they highlighted in their presentations would be aggravated or mitigated if there is a shift toward a more deglobalized trading system.

In response, Eswar Prasad mentioned policy frameworks, such as the dual circulation strategy, which use an indigenous innovation approach and suggested that these signal China's intent to move up the value-added chain and engage in technological upgrading. Prasad pointed out, however, that China is not self-reliant yet and needs American technology. He mentioned that China might be getting to a stage where economic and geopolitical tensions might affect its attempt to achieve domestic objectives.

Forbes then referred to the risks around inflation and fiscal policy that Fernanda Nechio highlighted in her presentation and asked whether a move toward a more deglobalized world would reduce the incentive to maintain fiscal discipline. Forbes inquired whether the ability to stabilize inflation in response to global shocks would be more or less difficult in a less globalized world.

In addition, Forbes posited that trade has been central to the China model. She asked Prasad what a more deglobalized world would mean for productivity and the reforms highlighted in his presentation. Would deglobalization slow down progress in that regard?

Finally, Forbes discussed what a shift toward friend-shoring or having two spheres of trade would mean for China, Brazil, and India. She asked whether this direction would hinder their development due to being less able to benefit from globalization, which has been key to the growth of many emerging markets, or instead improve their position by increasing their bargaining power as the largest world economies seek to expand their trade relationship with them.

Carlos Carvalho posited that a trend toward deglobalization would be less detrimental to Brazil's economy than a bipolar world given that Brazil is a very closed economy that exports a lot of commodities and trading with its important partners in a bipolar world would be more challenging.

Forbes also asked Acharya whether a shift to a more deglobalized world and less market discipline from trade would aggravate risks around market concentration in India, some of which were mentioned in his presentation.

Acharya responded to Forbes's questions by mentioning that if world economies become closed, there will be more concentration in sectors by construction, and he believes inflation will rise from nearshoring activity. Acharya expressed his view that integrating with the rest of the world is key for India to grow its share in manufacturing and become a slightly more exports-focused economy, as Rogoff mentioned in his comment on the three panel papers. Acharya stated that he expects India to remain a services hub but recognizes there are limits until the rest of the reforms (to improve education, skilling, and female labor force participation) are implemented.

Alessandro Rebucci commented on Acharya's focus on nonfinancial sector concentration and questioned whether this concentration might be rooted in concentration in the financial sector as well, and whether capital allocation in this sector might be driving concentration in both the nonfinancial and financial sectors. He also wondered whether there is a risk that the nonfinancial sector ends up regurgitating part of the financial system.

Acharya commented that in the past decade private banks have grown at the expense of public sector banks and mentioned the Reserve Bank of India (RBI) has prevented weak bank balance sheets from expanding, mentioning the RBI has adopted a "prompt corrective action" policy for weakly capitalized banks. He noted that even though more subtle, public sector banks are now becoming the largest asset management companies, with the biggest equity investments, making reference to the State Bank of India as the equivalent of Vanguard. Acharya mentioned that some of them are significant recipients of mutual fund flows and equity market allocations and noted the importance of researching the resulting implications for efficient capital allocation in the economy.

Tristan Reed commented on the potential assumptions in the markup estimation in Acharya's presentation and posited that including profit margins in the estimation, if not already included, would be more transparent.

Acharya responded that everything goes through if one limits the estimation to the traditional profit margin of companies, which he mentioned is what analysts generally do, and that he sought to do something better along the lines of state-of-the-art academic methods.

Blanchard raised the point that the interest rate at Brazil's central bank is set at 13.75 percent and the inflation rate is at about 4.7 percent, implying real interest rates of about 8 or 9 percent, highlighting that the minister

of finance of Brazil is not happy with that.[2] He wondered whether there are reasons to worry about the fiscal path and whether it is a cause or a result of the current fiscal situation in Brazil. Blanchard posed the question of whether the central bank is increasing rates too much, and what difficulties this may pose to the government.

Şebnem Kalemli-Özcan commented on Nechio's mention of Brazil's central bank grooming to the targeting regime, then losing credibility, and then regaining it quicker with the unanchoring index moving less. Kalemli-Özcan brought up how much the central bank of Brazil might care or not about the effects of the nominal exchange rate on inflation, suggesting that even though Nechio and Carvalho only show the second episode, the central bank was likely not interested in the exchange rate in the first episode, suggesting that they just looked at the unanchored expectations and increased the interest rate without mentioning any exchange rate to show capital loss issues. Kalemli-Özcan inquired whether the central bank did the same in the second episode but not in the first episode.

Nechio responded that more statistics, such as exchange rates, interest rates, and inflation, were included in the paper than in the presentation due to time limitations. She mentioned that the movements and effects had been very different in the previous episode in relation to the pandemic and that there was a sizable depreciation in the early stages of the pandemic, while the exchange rate did not move as much since the unanchoring of expectations. However, she emphasized that the same type of movements in effects leading during the period of unanchoring were seen in the first two episodes.

Nechio added that because of recent attacks to the law that granted central bank independence in Brazil in 2021, the authors decided to end their sample in January 2023.[3] She mentioned that, had they included the most currently available data, the degree of unanchoring would be at levels similar to 2015, when Brazil experienced the pursuit of a presidential impeachment and a downgrading. She noted that these attacks are making

2. Focus Economics, "Interest Rate in Brazil," https://www.focus-economics.com/country-indicator/brazil/interest-rate/; "Inflation in Brazil," https://www.focus-economics.com/country-indicator/brazil/inflation/; Marcela Ayres and Ricardo Brito, "Brazil Finance Minister Backs Lula on Rates but Is Focused on Economy," Reuters, February 9, 2023, Brazil finance minister backs Lula on rates but is focused on economy-source | Reuters.

3. International Monetary Fund, *Brazil: 2021 Article IV Consultation—Press Release; Staff Report; and Statement by the Executive Director for Brazil* (Washington: IMF, 2021), https://www.imf.org/-/media/Files/Publications/CR/2021/English/1BRAEA2021001.ashx.

it more challenging for the central bank to reduce inflation. On the topic of the interest rate level, Nechio mentioned that even though the central bank raised interest rates a significant amount, Brazil has rising and unanchored inflation expectations. She noted that many of their correlations are indicative of a relationship between the fiscal stance and monetary and inflation expectations, emphasizing that these are only correlations and that there is interest in researching this further in the future.

Carvalho echoed Nechio's point that Brazil has had high real interest rates over time and that this has been a debate for a long time. He suggested that looking at the record of inflation targeting could help understand whether rates have been excessively high. He mentioned that Brazil has rarely experienced below-target inflation, suggesting that interest rates have not been excessively high and that there is something structural about real interest rates in Brazil that is up for debate. Carvalho added that, for that reason, simple international comparisons of interest rate levels are not very useful.

Jason Furman brought up the distributional issues in China and his perception that even though inequality was increasing in the country, individuals at the bottom economic level were becoming better off and that now, with the Common Prosperity campaign, the rates of economic growth for those at the bottom have decreased. Furman interpreted this scenario as slowing economic growth for the whole Chinese population rather than increasing growth rates for the poorest, prompting the importance of thinking how this might unfold politically and economically.

In response, Prasad expressed his recognition of China's effectiveness at crafting narratives around its reform agenda when compared to the experience of other emerging economies. He brought up the point that internationalizing the renminbi for China meant opening up the capital account more cautiously, allowing the exchange rate to become more flexible with better financial markets, and better regulatory frameworks, which Prasad asserted would be better for China. He followed up by stating that the Common Prosperity campaign offers an avenue to change the narrative and offer support for reforms in light of concern in other emerging economies that economic and political elites will consume the benefits of the reforms. He also noted that China's efficiency in moving a lot of its people from the bottom end of the distribution to a better economic status resonates with the people of China.[4]

4. China Power, "Is China Succeeding at Eradicating Poverty?," October 23, 2020, https://chinapower.csis.org/poverty/.

Reed questioned the view of a decrease in the growth rate from 9 percent to 5 percent as catastrophic, referencing the point made in the Solow model that after capital accumulation the growth rate slows down. He questioned whether the decrease in China's growth rate should be seen as a disaster or as China becoming a middle-income or upper-middle-income country.

Prasad made the point that China has an economy that is over $18 trillion and that a growth rate of 5 to 6.5 percent in real terms adds significantly more to global GDP growth than any other country in the world.[5] He suggested that China may be too harshly judged by its historical standards rather than by international standards. However, he agreed with Rogoff that there are signs of a slower growth rate and pointed to whether and how much the government can change the allocation of resources to be more efficient and increase productivity growth.

In the context of the deglobalization debate, Goldberg pointed to Prasad's mention of capital flight from China and inquired where this capital would go, considering that it is likely no longer welcome in the United States or Europe. She mentioned that one possibility is that the capital go toward financing another of the BRI countries (Brazil, Russia, India), and she posed the question of whether it is safe to assume that capital would leave the country without knowing where it would go.

Prasad responded that capital flight is a concern from the private side and mentioned that it will be challenging for the People's Bank of China to move $100 billion elsewhere, considering the official capital flows China has had for several years before 2014 and 2015. However, Prasad stated that, as errors and omissions show, private capital is going out of the country and that it is unclear where the capital would go.

Goldberg also referred to the reversal in policy in 2021–2022 mentioned by Prasad and pointed out that it would make sense to seek to promote investment, given that there was no domestic demand due to China's zero-COVID-19 policy until recently, although this might be reversed again. Finally, Goldberg commented on Prasad's point that there has been little employment growth while the GDP growth rate has increased. She inquired whether, given China's demographic challenge of an aging population and low employment growth, promoting policies that increase labor productivity wouldn't be a sound strategy or whether there are missing considerations.

5. KPMG, "China Economic Monitor: Q1 2023," February 28, 2023, https://kpmg.com/cn/en/home/insights/2023/02/china-economic-monitor-q1-2023.html; China Power, "Unpacking China's GDP," January 12, 2023, https://chinapower.csis.org/tracker/china-gdp/.

Prasad asserted that there is an argument to generate growth in a country with a shrinking labor force through better investment but highlighted the importance of questioning whether the investment is efficient and financed by a financial system that is open to allocating domestic and international resources to the most productive uses. Prasad added that investment in China is being made largely by public firms and that the rate of return on investment does not look very good.

Prasad referenced Rogoff's work on the importance and risks of real estate investment to China's growth and posed the question of whether this would cause the system to come apart.[6] He mentioned that there are some developers and banks, especially the joint stock commercial banks and some of the rural credit cooperatives, that are very exposed, and he suggested that although it is very unlikely that this would cause a systemic event, there can be spillovers.

6. Kenneth S. Rogoff and Yuanchen Yang, "Peak China Housing," working paper 27697 (Cambridge, Mass.: National Bureau of Economic Research, 2020), https://www.nber.org/papers/w27697.

GEE HEE HONG
International Monetary Fund

DEBORAH LUCAS
Massachusetts Institute of Technology

COVID-19 Credit Policies around the World: Size, Scope, Costs, and Consequences

ABSTRACT Governments deployed credit policies on a historically unprecedented scale in response to the COVID-19 pandemic. We estimate the effective size of credit policies for seven large advanced economies in terms of the incremental resources provided to firms and households—a measure that allows aggregation across credit support, forbearance, and traditional fiscal policies but that does not appear in traditional government statistics. These estimates are used to reassess the absolute and relative size of different governments' policy interventions and to evaluate whether taking credit policies into account can help explain the cross section of macroeconomic outcomes. Incremental resources increase from an average 14.5 percent of 2020 GDP when only fiscal policies are considered to 22 percent of 2020 GDP when credit policies are also taken into account. Incorporating credit policies also reduces the cross-country variation in the total size of policy interventions. With regard to fiscal cost, fair value estimates for these credit support programs average 37 percent of principal, with wide variation depending on program features. We also discuss several related measurement issues, the financial regulatory changes that accommodated these programs, the pros and cons of the different types of credit policies, and how in principle budgetary costs should be calculated versus how governments account for credit policies in practice.

Conflict of Interest Disclosure: The International Monetary Fund reviewed this paper for disclosure of sensitive information. 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. The authors are not currently an officer, director, or board member of any organization with a financial or political interest in this paper.

Brookings Papers on Economic Activity, Spring 2023: 289–333 © 2024 The Brookings Institution.

Governments around the world deployed credit policies during the COVID-19 pandemic on a historically unprecedented scale. Those policies included loan guarantees, direct government lending, large-scale forbearance programs that allowed millions of households and businesses to temporarily stop making payments on certain debt obligations, and the loosening of regulatory restrictions on banks which encouraged them to participate in those programs. Advanced economies used both credit policies and traditional fiscal policies to a much greater extent than did emerging market and developing economies (EMDEs). Our quantitative analysis focuses on the credit policies of seven large, advanced countries (France, Germany, Italy, Japan, Spain, the United Kingdom, and the United States) that in dollar terms accounts for most credit policies that were implemented globally. The restriction of the analysis to those countries is for reasons of data availability, but the conceptual issues discussed are relevant to governments anywhere that are considering the use of credit policies as a complement to or substitute for traditional fiscal policies.

While a growing body of work studies various aspects of these credit policies, this paper appears to be the first to estimate the effective size of these policies in a way that allows aggregation across credit support, forbearance, and traditional fiscal policies, with the aim of producing statistics that can be used to explore the connections between these policies and various macroeconomic outcomes. Specifically, we introduce the idea of "incremental resources provided" as a unifying concept for sizing credit support, forbearance, and traditional fiscal programs. We operationalize that idea by equating size to principal take-up for government loan and loan guarantee programs, equating size with total avoided payments for forbearance programs, and equating size to reported expenditures for traditional fiscal policies. The size estimates for large-scale credit forbearance programs are an original contribution of this paper, while those for direct lending and guarantee programs are taken from Hong and Lucas (2023), and incremental fiscal spending is taken from the October 2021 release of the International Monetary Fund (IMF) database on fiscal policies in response to COVID-19.[1]

1. International Monetary Fund, "Database of Fiscal Policy Responses to COVID-19," https://www.imf.org/en/Topics/imf-and-covid19/Fiscal-Policies-Database-in-Response-to-COVID-19. The database was regularly updated from the onset of the pandemic to track governments' above-the-line and below-the-line activities that are categorized by the IMF as fiscal.

We use these size statistics, separately and in combination, to reassess governments' policy footprints during the pandemic, and also to explore whether taking large-scale credit and forbearance policies into account can better explain the realized cross-country differences in macroeconomic outcomes such as real GDP growth, private savings rates, and inflation. To briefly preview the main findings, we show that these credit policies significantly increased the resources in the pockets of firms and households, bringing the average share of incremental resources provided from 14.5 percent of 2020 GDP when only fiscal policies are considered to 22 percent when credit support and forbearance—as measured by take-up and missed payments—also are added in. Including those policies also paints a quite different picture of the relative aggressiveness of government policies across these countries. Whereas there is considerable variation in the use of traditional fiscal policies, the variation in resources provided as a share of GDP is much lower when credit and forbearance policies are also taken into account.

Furthermore, increases in savings rates between 2019 and 2020 across countries are highly correlated with the combined size of credit and fiscal policies, although forbearance policies have a negative correlation with savings rates. The much higher and more uniform levels of incremental resources provided by European and US governments, and the likelihood that a significant fraction of those resources were initially saved rather than spent, is consistent with pandemic fiscal and credit policies having significantly contributed to the sharp increase in subsequent inflation in all of these countries. In fact, the increase in private savings is highly correlated with inflation cross-sectionally. However, we find no direct correlation in the cross section between inflation rates and the size of these policies. Also suggestive of the importance of credit support and forbearance policies is the finding of a much stronger correlation between the cross section of GDP growth between 2020:Q3 and 2021:Q3 and our broad measure of incremental resources than with traditional fiscal spending alone.

There are a number of foundational conceptual issues related to credit and forbearance policies that are briefly discussed in section I. These include how the policies should be quantified, both for macroeconomic policy evaluation and for budgetary purposes, and why different approaches are required in each case; how to think about the transmission channels of credit policies and their interaction with fiscal and monetary policies; how to choose multipliers to convert raw measures of incremental resources into a more precise estimate of the contribution to aggregate demand from a given policy; the pros and cons of the different types of policies along

multiple dimensions; and how in principle budgetary costs should be calculated versus how governments account for credit policies in practice.

In section II we turn to the task of quantifying the incremental resources provided by each of the major credit support and loan forbearance programs and then aggregate the results in the two broad categories for each country. We also report estimates of the fiscal (i.e., upfront budgetary) cost for the direct lending and guarantee programs on a fair value basis, drawing on our recent estimates in Hong and Lucas (2023). Notably, the fiscal cost is typically much lower than the incremental resources provided, averaging 37 percent of the principal borrowed, because some or most of the funds will be repaid.[2] The estimated costs vary significantly across the programs depending on the generosity of program terms and the riskiness of the borrowers. We also describe the relaxation of regulatory rules that enabled banks to participate in these programs without incurring penalties. The important takeaway from these policies is that by lowering risk-weighted assets and raising consumer credit scores, they may have caused an overly rosy perception of the health of banks and the financial system. Section III reports on the macroeconomic results described earlier, and section IV concludes.

I. Effects of Credit Policies on Fiscal and Macroeconomic Outcomes: Conceptual Issues

A fundamental challenge for our analysis is one of measurement: how to best quantify credit policies so as to make them most comparable to traditional fiscal policies. Our preferred answer will differ depending on the outcome of interest. For example, we suggest that a credit program's effects on aggregate demand (i.e., stimulus effects) are best measured with an estimate of the incremental amount of cash the program puts into the pockets of households and firms, scaled by an appropriate multiplier that reflects factors such as the propensity of the recipients to spend the money rather than to save it. We also use an incremental cash approach to look for macroeconomic effects that are related to aggregate demand, such as GDP growth, saving, and inflation. However, we argue that a credit program's ex ante fiscal cost (or equivalently, its properly measured budgetary cost) should be evaluated in present value terms on a fair value basis in order

2. This excludes the US Paycheck Protection Program, which was effectively a grant program.

to make program cost estimates most comparable to those for traditional fiscal policies.

In addition to addressing measurement issues, as a prelude to the quantitative analysis it is useful to consider (1) the major types of credit policies introduced during the pandemic and their salient similarities to and differences from each other, (2) transmission channels for credit policy and the interaction of credit policies with monetary and fiscal policies, (3) the pros and cons of using alternative types of credit assistance and cash payments in terms of efficacy and cost effectiveness, and (4) how the cost of credit policies should be calculated versus how they are reflected in official government accounts. In this section we lightly touch on each of these issues and the reasons for the measurement choices made and point the interested reader to some of the related literature on these topics for more detailed discussions.

I.A. Types of Credit Policies and Comparisons between Them

A credit policy can be defined as a policy that affects the terms or availability of credit to households, firms, or subnational government entities. Here we classify the pandemic-era credit policies that we consider as falling into three broad categories.

The first category includes government loan guarantees and direct government lending. New loan guarantee programs for firms were the most prevalent form of credit assistance introduced in response to the pandemic. Loan guarantees and direct government loans typically are grouped together (e.g., by the IMF and the US Congressional Budget Office) because their benefits to borrowers and costs to the government are similar, holding the underlying loan and borrower characteristics constant. The similarities arise because through either form of support, the government absorbs some or all of the credit risk without being fully compensated for the cost of doing so. Furthermore, both types of assistance effectively put the government's credit rating in place of the borrower's rating. That allows firms and households that could not qualify for a traditional bank loan to gain access to credit, an often powerful extensive margin effect.

Whereas the economic effects of guaranteed loans and direct government lending are similar, there are several differences worth noting. The participation of private sector lenders in direct lending programs entails additional costs and benefits. It typically involves higher government cost because lenders have to be compensated for the costs and risks of their involvement, and that compensation may be in excess of what is required to induce participation. Part of that higher cost may be offset by efficiency

gains, such as when private lenders are better at screening borrowers or performing other administrative functions. For the government, the two types of credit support have very different effects on its balance sheet and the amount of government debt outstanding. Direct lending typically involves increasing government debt by the principal amount extended, and the loans are recorded on the asset side of the balance sheet. By contrast, credit guarantees do not entail any additional government borrowing, and they are generally treated as off-balance-sheet. Budgetary accounting for credit varies more widely across countries and in many cases is misleading, as discussed below.

A second category of credit policies that several governments introduced on an unprecedented scale during the pandemic were various types of payment holidays. These policies, which we refer to generically as forbearance policies but in some cases were called payment moratoriums, allowed borrowers with existing debt obligations to cease making scheduled payments without penalty for some period of time. As with guaranteed and direct lending, forbearance policies provided additional cash to households that would eventually need to be repaid. Effectively, forbearance is like making a loan to cover the payments coming due on existing debt. However, COVID-19 forbearance policies usually had a shorter maturity than guaranteed and direct loans, with the former typically ranging from a few months to a year and the latter typically lasting for three to ten years. The scale of funds provided relative to the underlying loan size is also different. With forbearance, the additional cash provided is the sum of missed payments over the life of the forbearance policy, whereas with a guaranteed or direct loan it is the entire principal amount. As discussed in section II.B, the size and incidence of the costs of forbearance varied considerably with the way the policies were structured. An important distinction is between forbearance programs that are imposed on private sector creditors and therefore entail hidden taxes and those whose costs are borne by the government.

A third category of credit policies that were important during the pandemic were changes to financial regulations—such as the rules governing when a loan is reported as nonperforming and how it is reported to credit registries. As explained in section II.C, regulatory requirements were relaxed to encourage program participation by both lenders and borrowers. Without these accommodations, the take-up of forbearance and guaranteed loans could have been much lower.

Some actions by central banks fall into the first category of direct lending or loan guarantees, but only some of those policies are included in this analysis. Newly created direct lending and guarantee programs that

were administered by central banks and backstopped by a fiscal authority were included. For example, we include the US Main Street Lending Program, which was administered by the Federal Reserve but whose losses were backstopped by the Treasury under the Coronavirus Aid, Relief, and Economic Security (CARES) Act. In principle we should also include any primary market purchases of private sector securities by central banks, but we did not obtain data to incorporate the effects of those purchases.

Other central bank policies would fall into a fourth category of credit policies that have more indirect effects on aggregate demand and more limited fiscal costs. As such, they would require a different sort of analysis than the one in this paper. These include purchases of private sector securities in the secondary market (which don't provide any new funding to firms) and the massive purchases of government securities that accommodated the fiscal response to the pandemic. In addition to outright securities purchases, central banks expanded their liquidity programs, reopening facilities created in response to the global financial crisis and in some instances adding new ones. For example, the European Central Bank expanded its targeted longer-term refinancing operations (TLTRO III) and introduced the Pandemic Emergency Purchase Programme (PEPP). The Federal Reserve reopened many of the facilities it created during the global financial crisis and added some new ones. While those liquidity facilities had the legal capacity to purchase trillions of dollars of securities, by design they took on very little uncompensated credit risk and didn't directly increase the supply of loanable funds to firms and households.

I.B. Transmission Channels and the Relation of Credit Policy to Fiscal and Monetary Policy

The basic transmission channels from all expansionary demand-side policies to the macroeconomy—credit, fiscal, or monetary—are broadly similar. All increase the resources available to households and firms and thereby encourage additional spending that in turn affects output, prices, asset values, and so forth.

The specifics of the mechanisms differ. Expansionary monetary policies make borrowing cheaper, encouraging investment and consumption (and vice versa for contractionary policies). When expansionary monetary policy is used to accommodate debt-funded government spending, that too can put upward pressure on prices. Expansionary fiscal policies, for example, extended unemployment benefits or tax cuts, lead to increased spending that puts upward pressure on prices and output. Expansionary credit policies lower the cost of credit and increase its availability, encouraging

increased consumption and investment that in turn influences macroeconomic outcomes.

It is important to emphasize that the intent of expansionary policies during COVID-19 was not to spur an increase in consumption during a period when people were encouraged or forced to stay home. Rather, the policies were intended to tide over households and businesses through a period of reduced economic activity, for instance, allowing firms to retain workers and remain going concerns and allowing households experiencing a temporary drop in income to continue covering fixed expenses such as mortgage or rental payments. Some described the policies as providing social insurance rather than as stimulating spending. However, the policies did increase aggregate demand relative to what it would have been in the absence of those policies, and hence it is reasonable to think of them as providing stimulus through aggregate demand effects.

Rather than according credit policy independent status, credit policies are generally treated as a subcategory of monetary, fiscal, and regulatory policies. While it is true that credit policy has aspects related to these other policy types and that it interacts with them, we hope something readers will take away from this analysis is that assessing the effects of credit policies accurately requires considering them separately in their own right.

In fact, credit policies have aspects that are fiscal, monetary, and regulatory. There is a fiscal element ex ante when a policy involves subsidies or ex post when it affects the future finances of the government, for example, when there are future payouts on loan guarantees. There is a monetary element when a credit policy originates from a central bank. There is a regulatory element when the policy affects administrative rules or legal constraints. As with fiscal policy, credit policies can be passive in that the effects arise from standing policies or active as when they are introduced in response to a shock. Also as with other government policies, their effects will vary over the business cycle and with other factors, making the magnitude and timing of their impacts difficult to predict.

In terms of policy interactions, credit policies affect the transmission mechanism of other types of policies, particularly monetary policy, by either blocking or amplifying their effects. The significance of those interactions can be illustrated by considering two historical examples. During the Great Recession, mortgage credit policy partially blocked the normal transmission mechanism for US monetary policy. Typically, one of the main ways that monetary policy easing helps households is via a mortgage refinancing channel. Households are able to reduce their monthly payment obligations by refinancing their mortgages at lower rates. However, because mortgage

credit policy prevented the refinancing of government-backed mortgages on homes whose value had fallen sharply or where minimum income requirements were no longer satisfied because of job loss, the volume of refinancing was much lower than would normally have been expected with such a large drop in interest rates (Remy, Lucas, and Moore 2011).

The interactive effect of credit policies with monetary policies during COVID-19 was the opposite. Both in the United States and internationally, credit policies amplified the effects of monetary policy easing. Loan guarantees, direct lending, forbearance, and relaxation of regulatory policies significantly reduced the barriers for households and firms to take advantage of the low level of real interest rates.

I.C. Credit Policy Multipliers

In order to evaluate the combined macroeconomic effects of credit policies and fiscal policies, a natural question is whether some scaling factors for our estimates of the additional cash made available are necessary to make the different policies more comparable to each other and to traditional fiscal expenditures? An example of such scaling is given by Lucas (2016), who uses a multiplier framework to translate an estimate of the incremental borrowing from US credit support programs during the Great Recession into stimulus estimates that are comparable to those for fiscal policy.

For loan guarantee and direct loan programs, we use take-up (the total principal amount borrowed) as the raw measure of incremental resources obtained by program participants. For forbearance programs, the estimated sum of skipped payments is the corresponding raw measure. The incremental fiscal expenditures are the government-reported cash costs of grants, tax cuts, and other discretionary spending, as reported by the IMF. Using take-up as the proxy for incremental cash made available by guarantee and direct lending programs presumes that the loans would not otherwise have been made, that is, that there was no crowding out of private sector credit. For small and medium-sized enterprise (SME) programs, which account for a large portion of total take-up, it is reasonable to assume that there would have been limited credit availability without government support, or that rates would have been so high as to discourage most borrowing.[3]

3. Lucas (2016) assumed variation across US government credit programs in incremental resources made available during the Great Recession, with borrowing under programs like student loans judged to be largely incremental and borrowing under government mortgage programs less incremental.

For the programs serving larger firms, presumably some would have obtained funding in the absence of government support, and take-up probably overstates incremental resources provided.

Choosing multipliers to convert these raw measures of incremental resources into an estimate of their contribution to aggregate demand is especially challenging for the pandemic period. As noted earlier, the intent of assistance programs during COVID-19 was quite different than that of stimulus programs during past recessions. Rather than to bolster aggregate demand, the assistance was intended to tide over individuals and firms through a period of drastically reduced incomes and heightened uncertainty so as to be able to cover basic needs and minimize displacement. Auerbach and others (2021) address these challenges to multiplier estimation in the context of fiscal policy and emphasize that multipliers were much higher in places with less strict lockdown orders.

Given the difficulty of inferring multipliers for this period, in the analysis that follows we have not applied any scaling factors to our raw estimates of the incremental cash provided by credit assistance programs or fiscal expenditures. However, it is worth considering, at least qualitatively, the extent to which the incremental cash provided via credit programs could be expected to have increased aggregate demand and the reasons the effects may be larger in some programs than in others.

GUARANTEED AND DIRECT LENDING PROGRAMS There are several reasons to expect that a significant share of the funds borrowed were retained for precautionary purposes rather than immediately spent: demand was subdued by lockdowns, caution and economic uncertainty were high, and the subsidies in some of these programs were large enough to make it worthwhile to borrow even if there was no immediate use for the funds. To the extent borrowed funds were spent, for the same reasons the size of subsequent rounds of spending that were triggered may have been much smaller than in a typical recession. These considerations point to smaller multipliers than in prior recessions.

Anecdotal evidence about how the borrowed funds were used supports a higher-than-normal propensity to save the borrowed funds. For instance, according to the British Business Bank, 23 percent of SMEs had spent all their facilities and 19 percent had not spent any by 2020:Q3 (British Business Bank 2021). The gross savings statistics reported in section III also suggest a high propensity to save during this period.

The question remains of whether the multipliers applied to take-up amounts should be lower or higher than those that are on average attached to traditional fiscal expenditures. Lucas (2016) noted that credit multipliers

will tend to be higher than fiscal multipliers, all else equal, because it is costly to borrow and people will refrain from doing so unless they intend to spend the money. However, for the reasons just mentioned, during the pandemic it was relatively low-cost to borrow and the demand for precautionary stockpiles of cash was likely to be elevated. Lacking a basis for adjusting the relative multipliers on credit and traditional fiscal assistance, we concluded that a neutral choice was to set them to be the same and equal to one.

FORBEARANCE Forbearance policies provided additional cash to affected firms and households in the amount of the skipped payments. The additional cash on hand provided funds that would have been otherwise unavailable for spending or for saving. Because most or all of the money had to be repaid eventually, and often with additional accrued interest, the policies should have had minor wealth effects. The primary transmission channel is therefore most likely to be via the effective relaxation of borrowing constraints. These effects are akin to other forms of credit support that operated primarily through increased borrowing on the extensive margin.

The questions of (1) how much of the missed payments were saved rather than spent and (2) whether the propensity to spend out of the additional funds was similar to the propensity to spend out of incremental borrowing from guaranteed or direct lending programs or from fiscal policies can't be answered with the aggregated program data we have available. It seems likely that the propensity to spend varied depending on whether a program was for households or firms and on other borrower and program characteristics. In programs where forbearance was automatic or close to it, more of the funds were likely to have been saved than in programs where some demonstration of hardship was required. There may also have been a lower average propensity to spend than for guaranteed and direct loans, which even at subsidized interest rates entailed higher costs, more effort, and more selectivity.

A further consideration regarding the spending impact of forbearance programs is whether there was a partially offsetting effect from creditors that may have reduced spending or lending to other borrowers, especially for those where forbearance arose from an uncompensated mandate.[4] Given the regulatory forbearance discussed below to accommodate these programs and the generally strong capital ratios of banks, we expect these offsetting effects to have been modest.

4. We thank Alan Auerbach for pointing this out.

I.D. Policy Trade-offs between Credit, Forbearance, and Fiscal Policies

We will see that credit support and forbearance policies appear to have been used as both substitutes and complements to traditional fiscal policies during the pandemic. What criteria should have guided the policy mix? An investigation into these questions is important for informing future policy choices. While an in-depth analysis of the quality of choices made during the pandemic is beyond the scope of this paper, it is useful to evaluate the trade-offs in general terms. Here we briefly discuss some of these considerations, drawing on Lucas (2020).

In choosing between different types of policy responses, important considerations include efficacy, target efficiency, transparency, the costs of financial distress to households and firms, and government cost. In terms of target efficiency and cost transparency, traditional fiscal policies (i.e., cash transfers, tax cuts, and in-kind assistance) will often dominate credit. Such assistance usually reaches the targeted recipients, and it can be distributed relatively quickly and through existing administrative structures. With some notable exceptions, it is transparent who benefits and how much they receive. It doesn't create an overhang of future obligations on recipients. Importantly, cash can help poorer households and proprietors who are less connected to credit markets. For instance, landlords benefit from mortgage forbearance and renters benefit from cash rental assistance. Cash transfers also have the advantage of greater cost transparency.

The main disadvantage of traditional fiscal measures is that they are expensive, costing dollar for dollar the amount of purchasing power provided. That can make it infeasible to provide a meaningful level of assistance on a large scale. When the goal is to offset a big but probably temporary drop in income or to mitigate a sharp contraction in the availability of credit from private financial institutions, credit support can be an attractive alternative.

The cost of credit support to the government can be quite low relative to the amount of funds that are made available, making it a potentially cost-effective way to get money into the hands of people and businesses during times of crisis when liquidity is scarce. That's because most obligations incurred are likely to be eventually repaid. Further, the obligation to repay makes it more credible that those applying for funds truly need the money.

However, the stated advantages of credit presume a program that is well designed so as to reach borrowers who are likely to have the capacity to repay and to screen out those who are not targets for assistance. A striking example of a credit program that had none of these potential advantages was the US Paycheck Protection Program, whose design has been

criticized as a poorly disguised grant program that lacked transparency and target efficiency. A further consideration is that defaults are costly to all parties involved—the defaulting borrowers, lenders, and the government agencies administering the programs. When those costs are taken into account, a loan program with a high expected default rate is likely to be more expensive than providing a similar amount of cash assistance.

Already high government debt levels and restrictions on deficit spending may discourage some countries from increasing traditional fiscal assistance. Credit, for better or worse, may be a legal way to avoid constraints on traditional spending. Large cash transfers also are more likely to raise issues of fairness than debt obligations that have to be repaid.

Among credit policies, a well-crafted forbearance program can have several advantages under certain circumstances. It can get money out the door quickly with little new bureaucracy, and it can be targeted fairly tightly, such as at residential mortgages with some maximum balance. When the government already bears most of the credit risk for loans to borrowers that are targets for assistance, the incremental cost and risk of forbearance is likely to be small. This is particularly true in the United States, where, through its mortgage, student loan, small business, agricultural, emergency, and other lending programs, the federal government is the largest provider of credit to US households. Many of those programs already allowed for forbearance during an emergency without additional legislative action, which made it possible to quickly put the policies into motion. Beyond the direct benefits of reducing cash needs, forbearance policies can avoid longer-term economic damages to employment and access to credit markets by protecting household credit scores.

Forbearance can also have a downside, as discussed at greater length in section II.B. It can only help the limited number of households and firms with existing debt, and this group is likely to be relatively well-off. When it is combined with partial forgiveness of principal or interest, its costs can be much higher and more akin to cash grants, and those policy changes may circumvent the discipline of recognition in the budget process. Forbearance can be target-inefficient when it is made available without consideration of need. The US student loan moratorium is an example that has all of these drawbacks. Forbearance also can be used as a way to avoid recognizing fiscal costs by imposing costs on the private sector. For example, in the case of US mortgage forbearance, servicers were compelled to continue to pay the holders of mortgage-backed securities even on mortgages on which payments had stopped. Bank loans subject to loan moratoria in Europe also presumably involved uncompensated costs to banks.

New credit guarantee or direct lending programs are harder and slower to put in motion than forbearance policies, often requiring new legislative authority, obtaining special expertise from outside government, delays associated with vetting borrowers, and so on. Nevertheless, they can reach firms and households that don't have existing debt, which are often more in need of funds than those that already have a connection to the credit market. The estimates of government cost developed in Hong and Lucas (2023) and reported below suggest that most of the credit support programs introduced during the pandemic had a significantly lower government cost than had the same amount of funds been provided through traditional fiscal spending.

I.E. Estimating and Accounting for the Fiscal Cost of Government Credit Support

What is the most policy-relevant way to assess and budget for the fiscal cost of credit support? And how does the answer compare to how it is budgeted for in practice? We discuss these issues at length and apply them to the COVID-19 guarantee programs in Hong and Lucas (2023). The main takeaways are briefly summarized here; the interested reader is referred to that paper and references therein.

The policy relevance of budget estimates depends on the purposes for which they're used. Arguably, the most important purpose is as an input to help policymakers understand the trade-offs between competing uses of fiscal resources, and to do so upfront, at the point in time authorizing legislation is being drafted and choices between competing actions are being made. Cost estimates also feed into tabulations of aggregate surpluses and deficits, which in turn provide policymakers with a signal about the stance of fiscal policy. The question then is, What basis of upfront cost estimation and accounting for credit support puts it on a level playing field with other types of fiscal policies? This is sometimes referred to as seeking "grant equivalence," whereby the reported cost of credit assistance can be interpreted as the size of a grant that would have the same fiscal cost to the government.

A fair value approach, which equates the estimated cost of credit support with its market value (or an estimate thereof when comparable and reliable market prices are not available), is our choice for achieving grant equivalence. An example that illustrates why a fair value approach achieves that goal is to note that the fair value of a loan guarantee is the price a financial institution would charge for providing an identical guarantee. Therefore, a government could offer grants to program participants to cover the price of loan guarantees purchased from banks or it could buy the guarantees on behalf of the participants. The outlay would be the same in either case.

Notice too that the price charged by a private sector guarantor would reflect the present value of the future uncertain cash flows associated with the guarantee. The discount rates that market participants would implicitly or explicitly assign to those cash flows would compensate for time value and also include a risk premium.

As the previous example foreshadows, operationalizing the idea of fair value cost in a budgetary context requires accounting for credit on an accrual (i.e., present value) rather than a cash basis. What distinguishes fair value accruals from alternative accrual approaches is that the discount rate reflects the risk of the cash flows—the discount rate is "risk adjusted." Risk adjustment recognizes that taxpayers and other government stakeholders ultimately bear the associated credit risk and effectively are shareholders in risky government investments. Online appendix II outlines the specific approach we used for the subsidy cost estimates for the COVID-19-era direct loans and loan guarantees we evaluated. We have not attempted to estimate the fiscal cost of forbearance programs, although we discuss the costs qualitatively below.

In practice, most of the countries covered in this analysis, as well as most countries that are not included, do not record any upfront cost of credit programs in their budgets. Instead, they report credit assistance "below the line," which means there is often no immediate budgetary impact from a new credit assistance program. Losses are only reflected when they eventually are realized. An exception is the United States, which reports credit subsidies on an accrual basis in the federal budget. However, those accruals are required by law to be calculated by discounting expected cash flows at government discount rates. That legal restriction causes the reported costs to be systematically lower than fair value estimates. The Congressional Budget Office has supported fair value cost estimates for credit and often provides supplemental estimates for major credit programs on a fair value basis.

For the purpose of evaluating the macroeconomic effects of credit policy during COVID-19, we follow Lucas (2016) in treating fiscal cost as much less important than the incremental amounts of funds made available. The idea is that the expansionary effects of these policies during episodes of severe economic stress are primarily on the extensive margin of credit provision, rather than on the intensive one where the size of the subsidies is relevant. However, estimates of fiscal costs are important for comparisons of bang for the buck of different policies: the amount of stimulus provided relative to its fiscal cost. We provide some rough comparisons across countries of bang for the buck for the loan guarantee and direct lending programs.

II. Pandemic Credit Policies: Evaluation of Size and Fiscal Costs

In this section we seek to provide a quantitative answer to the deceptively simple question: How much incremental cash did households and businesses obtain via credit support and forbearance programs in different countries in aggregate? As discussed in section I, the answer to this question is a necessary input for analyzing and comparing the broader macroeconomic effects of these policies in a way that is fairly parallel to analyses of the effects of traditional fiscal spending. For the major credit guarantee and direct lending programs, we also report estimates of their fiscal cost (taken from our previous work), which when compared to the incremental cash provided provide bang for the buck estimates for these programs.

In section II.C we also briefly discuss the relaxation of regulatory requirements and other rule changes affecting banks that accommodated the large-scale usage of credit guarantee and forbearance policies. Notably, those policies may have contributed to overly optimistic perceptions about the health of the financial system and of household finances during this period.

We consider credit support in the form of loan guarantee and direct lending programs separately from the assistance provided by forbearance policies. The many differences between these policy types raise somewhat different conceptual issues and necessitate applying different but related measurement approaches. Our estimates draw on, and extrapolate from, data obtained from a variety of sources including official reports and statistical sources, academic and policy papers, and our own recent work on some of these programs.

II.A. Credit Guarantee and Direct Lending Programs for Firms

An intriguing database released by the IMF in 2021 provided an international comparison across countries of pandemic fiscal and credit policies. Specifically, it compared incremental traditional fiscal spending related to COVID-19 with the program "envelope" for below-the-line (off-budget) measures. Traditional fiscal spending includes both health- and non-health-related items.[5] The envelope of below-the-line measures includes the amounts

5. We find that the IMF's measure of incremental spending is, for some countries, similar to, and overall, highly correlated with the excess of fiscal expenditures in 2020 and 2021 over those in 2019.

that were legally authorized for new credit support and equity purchase programs. We equate the envelope with new credit support policies because equity purchases comprise a small share of the totals.[6] Those data suggest the potential importance of credit policies during COVID-19. It reveals the wide variation across countries in the types of policies introduced and the much less aggressive policy responses of emerging markets. The graphs summarizing the IMF findings are reproduced in figure 1.

While the IMF data are suggestive of the potential importance of credit policies during this episode, the credit envelope has only an indirect connection to macroeconomic and fiscal outcomes. In most cases, the envelope is much larger than the amount of credit extended under the programs—the program take-up. Furthermore, the relation between take-up and envelope is a noisy one. The envelopes reported by the IMF also do not include any of the funds made available through the large-scale forbearance programs discussed in section I.

To answer questions about how much money these credit programs made available to borrowers, the fiscal costs, and the stimulus provided, take-up is the more relevant starting point and the one we focus on here. The statistics reported on take-up and cost in this section are largely drawn from our recent paper, Hong and Lucas (2023). In that study, we provided estimates of take-up and fiscal cost on a fair value basis for the major credit support programs for firms that were implemented in seven countries, including the five largest economies in Europe (France, Germany, Spain, Italy, and the United Kingdom), Japan, and the United States.[7] Overall, these programs covered more than 90 percent of the credit support programs for firms that were introduced around the world during the pandemic according to the IMF. Most of these programs were launched following the lockdown orders in March 2020, in anticipation of a wave of firm defaults and a drop in income and liquidity.

Figure 2 compares the total program envelopes with realized take-up, both in dollar terms and as a share of 2020 GDP. In dollar terms, the United States had the largest envelope of about $1.4 trillion. When normalized by GDP, however, several other countries had larger envelopes. The total take-up of credit support by firms in the seven advanced economies we

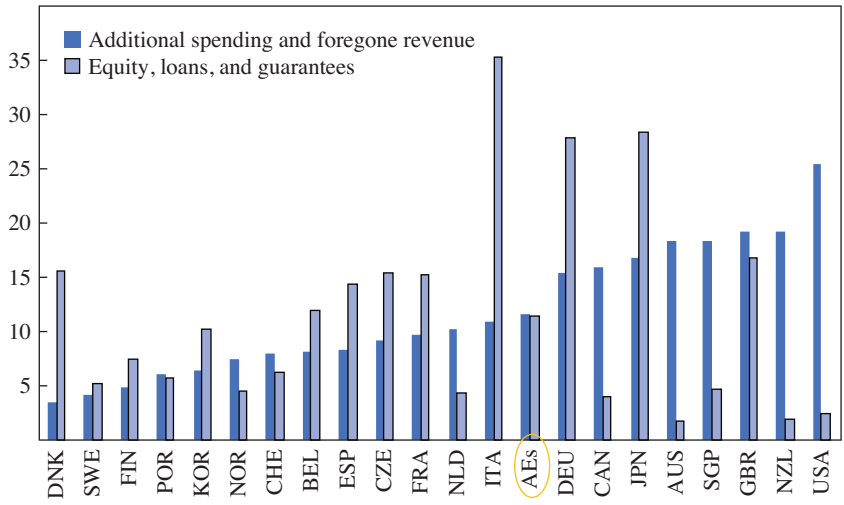
6. The IMF categorizes all of the policies as fiscal and breaks them into these two subcategories: “additional spending and forgone revenue” and “equity, loans, and guarantees.”

7. For details of individual programs and countries, see the annex of Hong and Lucas (2023).

Figure 1. Discretionary Fiscal Response to the COVID-19 Crisis in Selected Economies

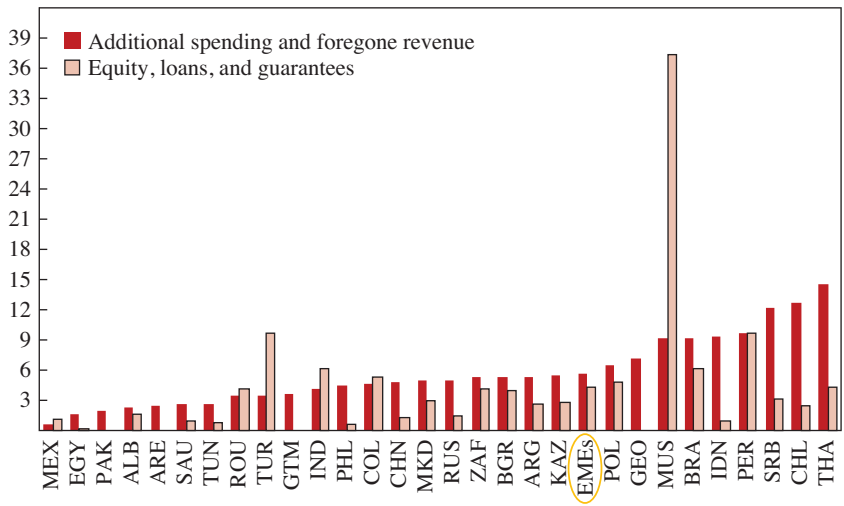
Panel A: Authorized fiscal support for selected advanced economies

Percent of 2020 GDP



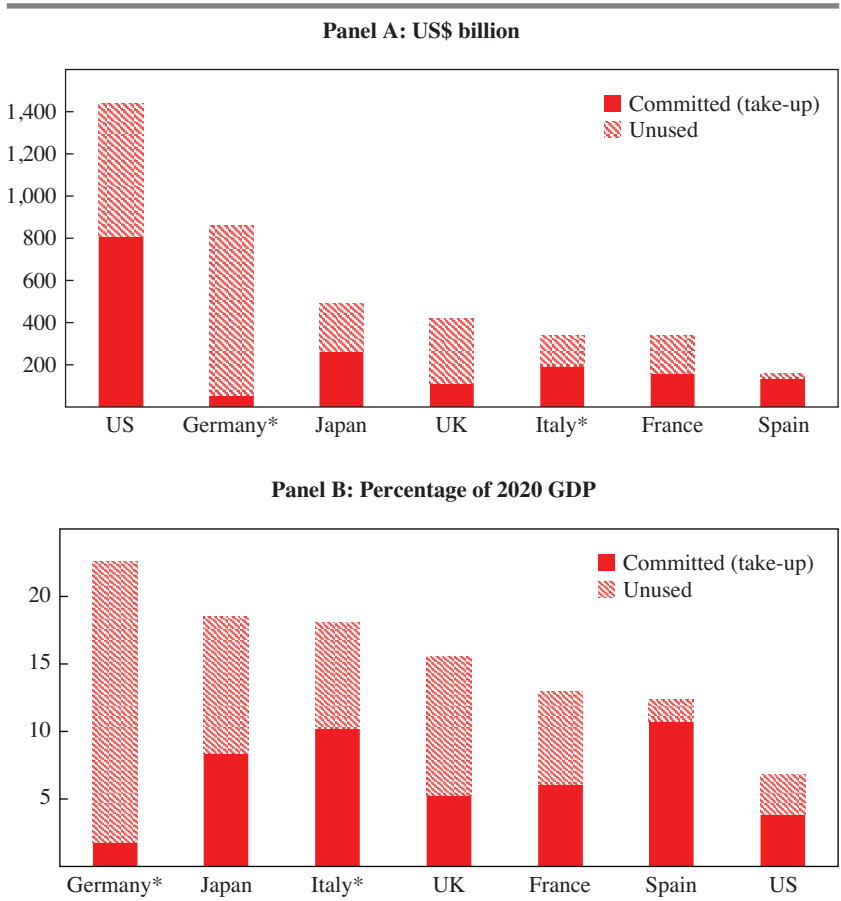
Panel B: Authorized fiscal support for selected emerging market economies

Percent of 2020 GDP



Source: Authors' tabulations using the IMF Database of Country Fiscal Measures in Response to the COVID-19 Pandemic. Panel A was adapted from Hong and Lucas (2023).

Figure 2. Credit Support Programs during COVID-19: Committed versus Unused Envelope



Source: Reproduced from Hong and Lucas (2023); estimates are based on official sources for each country, IMF Fiscal Monitor Database of Country Fiscal Measures in Response to the COVID-19 Pandemic, and Anderson, Papadia, and Véron (2021).

Note: The following programs are included in each country's total envelope: Italy—"Cura Italia" guarantees and SACE Garantia Italia; United Kingdom—BBLS, CBILS, CLBILS, and the Bank of England's COVID corporate financing facility (CCFF); Spain—ICO loan guarantees; France—Prêt Garanti par l'État; Germany—KfW loans for small and large loans and a part of the increase in the economic stabilization fund (WSF) (400 billion euros) to provide additional state guarantees for firms; Japan—Safety Net No. 4 and No. 5; United States—Paycheck Protection Program, Main Street Lending Program, Credit Support for Airlines and Critical Industries. Last observations were in January 2022, except for Japan (January 2021).

*Some guarantee programs announced by Germany and Italy have uncapped legal limits to provide funds. The bars in the figure show the actual injections to the guarantee funds rather than the cap on total guarantees by the government.

study exceeded $1.7 trillion.[8] Using a simple average across countries, the average take-up is about 6 percent of 2020 GDP.

The wide variation in observed take-up rates has a variety of possible explanations. Countries that experienced larger economic shocks from the pandemic may have had higher take-up rates, other things equal. At the same time, the attractiveness of program terms, bottlenecks in financial intermediaries in assessing loans, and the availability of noncredit support programs are at play (Anderson, Papadia, and Véron 2021). In Spain, for instance, the greater recourse to guaranteed loans can be partly attributed to the lower availability of alternative fiscal relief measures for corporations (e.g., direct grants of state aid). This is consistent with the general pattern we find in section III of what appears to be substitution between credit and noncredit assistance. In France, the higher take-up of guaranteed loans may reflect the favorable pricing offered to borrowers through grace periods and concessional rates, especially during the first year of the loan. In Germany, the relatively limited use of such loans may be due to (1) lower financing needs of firms compared with other countries owing to a less stringent lockdown and firms' greater use of a combination of other policy measures that supported corporate balance sheets, including direct grants and tax deferrals and short-time working allowances; (2) less favorable lending terms (e.g., higher rates, a prohibition on distributing dividends, and limits on the remuneration of managers); and (3) some supply-side bottlenecks related to the risk assessment required for large loans. On the loan supply side, operational bottlenecks and an overwhelming number of loan applications may have impeded the take-up at the initial stage of the programs. Core and De Marco (2021) look at the different levels of information technology used to process online applications by banks in Italy and emphasize the role of information technology in processing high volumes of online applications and disbursements. Over time, the take-up in Italy continued to increase, notwithstanding the existence of a large-scale debt moratorium scheme.

The terms of these credit programs varied along a variety of dimensions. To give a fairly typical example, consider the UK Coronavirus Business Interruption Loan Scheme (CBILS), a guarantee program directed at small and medium-sized businesses.[9] It provided loans ranging from GBP 50,000 to

8. Most credit programs reported take-up through January 2022, except for Japan (January 2021).

9. Gov.UK, "HM Treasury Coronavirus (COVID-19) Business Loan Scheme Statistics," https://www.gov.uk/government/collections/hm-treasury-coronavirus-covid-19-business-loan-scheme-statistics.

GBP 5 million, with the available amounts depending on firm characteristics. Loan maturities ranged from three months to six years, and a variety of loan types were eligible to be guaranteed. Collateral was not required on most loans. The program provided an 80 percent guarantee, with 20 percent of losses absorbed by the private sector lenders. Lenders were charged an annual guarantee fee of 75 basis points. The government paid the first twelve months of interest and fees, up to a maximum of GBP 800,000. Loan pricing was at the discretion of lenders, but lenders had to demonstrate that the net financial advantage of the guarantee was passed through to the borrower.[10] The program wound up authorizing 110,000 loans totaling GBP 26.4 billion.

Structurally, there were common elements across most credit guarantee programs. First, a program specified the characteristics of target beneficiaries. In some programs, companies of all sizes were eligible (e.g., Germany's KfW, the French state-guaranteed loan PGE), while in others, only companies of certain sizes could participate (e.g., CBILS versus the Coronavirus Large Business Interruption Loan Scheme in the United Kingdom). Second, program rules specify guarantee coverage, that is, the share of losses absorbed by the government in the event of a default. In the programs that we assess, the guarantee coverage ranged from 70 percent to 100 percent.[11] Third, program rules restrict other terms, such as how interest rates are set, loan maturities, and eligible loan types (e.g., term or asset-backed loans), and guarantee and other fees or premiums. In some programs, interest rates were fixed by the government (e.g., the United Kingdom's Bounce Back Loan Scheme, BBLS), and in others, lenders were permitted to set the rate but with a cap (Spain) or subject to benefit pass-through (e.g., CBILS and PGE). For countries in the European Union, guarantee fees are prescribed in the EU temporary framework based on maturity and firm size.[12] Finally, loan sizes were typically limited in absolute amounts. For countries subject to the EU temporary framework, loan size was subject to a ceiling where the total amount should not exceed (1) double the annual wage bill of the

10. Although such requirements might not have been strictly adhered to in all cases, the considerable regulatory oversight of the banking system and the adverse consequences of being found to have violated the rules provide incentives for compliance. Anecdotal evidence suggests such rules were taken seriously.

11. The guarantees were typically pari passu, meaning that losses were shared proportionally between lenders and the government.

12. Under the EU temporary framework, guarantee fees (premiums) range from 25 basis points to 200 basis points, increasing progressively in line with the duration of the guaranteed loan and firm size.

beneficiary for 2019 or for the last year available, or (2) 25 percent of the beneficiary's total turnover in 2019. Exceptions could be made if there was appropriate justification and self-certification by the beneficiary of its liquidity needs. Several countries included additional provisions to allow loans to be more generous, such as a one- or two-year waiver of principal payments (pre-amortization). In some countries like the United Kingdom, interest payments were also paid by the government for one or more years.

To make the loans more accessible to certain target borrowers, several countries introduced programs with 100 percent government guarantees and relatively generous terms. There were five programs of this type that are included in the statistics here: the US PPP, Germany KfW's Instant Loan, the UK BBLS, Japan's Safety Nets for Financing Guarantee Nos. 4 and 5, and Italy's Fondo di garanzia. Compared to other programs under the same umbrella but with partial guarantees, these full-guarantee schemes had (1) quicker disbursement and minimal credit risk assessment, (2) longer loan maturities, and (3) lower maximum loan amounts. In some programs, terms were liberalized over time with extensions of loan maturity, extensions of program end dates, or increases in the envelope. The reported take-up and envelope sizes reflect the most recent information on those totals.

FISCAL COSTS OF CREDIT SUPPORT PROGRAMS We equate the fiscal cost of credit support programs to the fair value of the assistance granted, as estimated in Hong and Lucas (2023). The fair value estimates represent the net present value of projected cash flows to and from the government over the life of the underlying loans, approximately as of the point in time when the loans were originated (see online appendix II for a stylized example). Discount rates are inferred from quoted or observed market rates, adjusted using fair value principles for rate determination. Subsidy rates are defined as the ratio of fiscal cost to loan principal at origination. As discussed in section I.E, the aim is to produce cost estimates for credit support that are grant equivalent, meaning that the cost to the government is equivalent to that of providing program beneficiaries with an upfront cash grant that would allow them to obtain credit on the same terms from a private sector financial institution.

We estimated a total fiscal cost of $330 billion ($1.13 trillion including the US PPP). That is relative to an estimated total take-up of $873 billion ($1.76 trillion including PPP). Dividing total fiscal cost by total take-up, the average subsidy rate is 37 percent (67 percent including PPP). The subsidy rate varies widely across programs as a function of program design choices such as the riskiness of target borrowers, the size of rate concessions, loan maturity, fees, and other program features.

In terms of cost efficiency or bang for the buck, a subsidy rate of 37 percent represents significant savings over the 100 percent subsidy associated with traditional fiscal spending. Whether a particular credit program is truly more cost effective than a fiscal action aimed at the same outcome will, of course, depend on additional factors such as whether the targeted recipients are the ones to take advantage of the credit program, whether the funds are put to their intended uses, and whether the multipliers translating take-up to ultimate spending growth are in fact similar to the multipliers on fiscal spending.

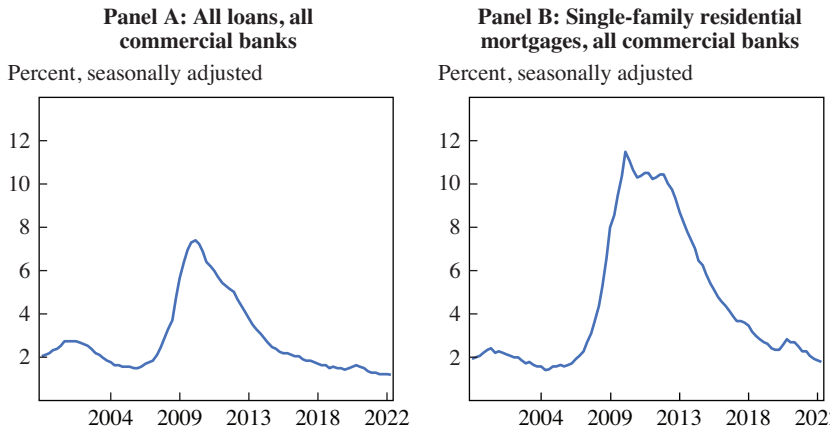
II.B. Forbearance Policies

Many countries adopted debt moratoria and other types of payment holidays on debt (collectively referred to here as forbearance policies) that allowed millions of households and businesses to postpone payments on existing financial obligations for periods ranging from a few months to several years. Most forbearance policies were initiated in spring of 2020 and ended later that year or in 2021. For households internationally, the largest source of payment relief came from mortgage forbearance, with rent, auto loan, and student loan payment stoppages also providing significant assistance in some countries. Forbearance measures for nonfinancial corporations (NFCs) typically were aimed at SMEs with bank loans.

Programs offering rent relief were not included in our tabulations of forbearance. For rent relief that was funded by a national government, the expenditure should be captured in incremental fiscal spending. For moratoria imposed on landlords as a mandate without compensation, funding was effectively via an un-booked tax that wouldn't be captured in any of the data that are available. To the extent that the increased propensity to spend by renters is offset by a reduced propensity by landlords, the omission of those programs should not distort inferences about macroeconomic effects. However, if renters have a higher marginal propensity to consume than landlords, omitting such mandates understates the stimulus effects of government policies.

The goal of forbearance policies was to avoid defaults by the many households and businesses that experienced sharp drops in income that were expected to be temporary. Avoided defaults prevented costly disruptions to lives and businesses. It stopped people from losing their homes, helped firms to remain as going concerns and retain employees, and shored up banks' regulatory capital. As others have noted, the extensive use of forbearance is one explanation for the historically low default rates that have continued through the present day. For example, figure 3 illustrates the

Figure 3. Default Rates on US Loans and Mortgages



Source: FRED Economic Data.

historically low default rates in the United States during the pandemic and in its aftermath. Forbearance policies also carry costs and risks. To highlight a few of them, there is the potential to leave households with unaffordable levels of future debt and to create zombie firms whose eventual insolvency is more costly and potentially more destabilizing than had they been promptly liquidated (see also section I.D). To the extent the loans are treated as performing, the health of the financial system appears better than it is.

Forbearance policies—particularly those that were made widely available such as the moratorium on US student loan payments—also allowed many non-distressed borrowers who could have serviced their loans to stop making payments. The excess funds were then available to save or spend. For example, Albuquerque and Varadi (2022) present evidence for the United Kingdom that the mortgage holiday increased consumption among poorer households and savings among wealthier ones. Importantly, because most missed payments would not come due for several years, the effects of forbearance on savings and aggregate demand could extend well beyond the date that programs officially ended.

In the analysis here, we do not try to identify distressed versus non-distressed recipients of forbearance assistance or its target efficiency. Rather, the more modest goal is to estimate the total amount of incremental funds provided by missed payments under these programs in different countries and regions. The estimates are new to this paper, and we haven't found other studies that try to estimate these quantities in aggregate. Later in this

Table 1. Total Missed Loan Payments due to COVID-19 Forbearance Policies

Country	*Forbearance (US$ billions)*	*Forbearance (% GDP)*
DEU	3.5	0.1
ESP	46.3	3.6
FRA	63.2	2.4
ITA	38.7	2.1
JPN	0	0.0
UK	31.65	1.2
USA	117	0.6

Source: Authors' estimates.

section we add those missed payments to the funds obtained through take-up of credit support programs and from traditional fiscal policies to provide a comprehensive measure of incremental funds going to households and businesses through all major types of government policies during the pandemic.

The results of our investigation into incremental resources provided by large-scale credit forbearance programs, as measured by estimates of total missed loan payments due to forbearance policies, are summarized in table 1. There are several important caveats. The first is that it is possible that there were significant forbearance policies that are left out of these tabulations. Second is that the estimates rely on extrapolations and approximations rather than on direct observations of the volume of missed payments. Third is that there is no offset for any reduction in lender resources when the programs involved uncompensated mandates on the private sector. Nevertheless, they represent our best estimates and are used as inputs into the macroeconomic analysis of section III. Readers primarily interested in those results rather than the details of how these totals were arrived at can safely skip to that section.

ESTIMATING THE SIZE OF INCREMENTAL RESOURCES FROM FORBEARANCE Data on forbearance programs are (even) scarcer than for credit support programs. Ideally, banking regulators or government agencies would monitor government forbearance programs and keep track of the size and timing of missed payments and the characteristics of program beneficiaries. In addition, keeping track of longer-run performance statistics would provide insights into the costs and benefits of such assistance that could inform future policy choices.

We have been able to obtain only limited data on COVID-19 forbearance policies, and only for some of the larger programs. The difficulty of obtaining data is partly due to the variety of channels through which the programs

came about. Some of the programs were authorized through legislation while others came about by administrative decree or were voluntarily instituted by private entities. Some programs that arose from local directives might not be included in national reports. Some programs mandated that lenders participate while others gave lenders more discretion. In many instances the programs were extended beyond their originally announced end date and we are unsure about the ultimate end date or how strictly it was held to. Information about how the missed payments would be treated is also not readily available. For most programs we don't know whether, and at what rate, interest accrued on missed payments or what the scheduled timing was for when missed payments would be recouped.

Despite these many uncertainties, our estimates suggest that forbearance is likely to have provided a significant amount of additional cash to households and firms in some countries. Although there was likely to be some offset caused by the reduced resources of lenders for whom some of the policies were an unfunded mandate, because banks were well capitalized and liquidity was plentiful the offset is expected to be small. Quantifying the additional resources allows us to consider the aggregate effects of those policies.

Our estimates of the total value of missed payments arising from a given policy rely on combining data from various reports and press releases with assumptions to fill in any missing information about borrower types, loan maturities, program duration, and so on. For most major programs in developed countries, we have information on the principal value of loans that received relief and the types of eligible loans. We can estimate the payment reduction per dollar of principal per period (e.g., monthly) based on typical loan maturity and interest rate by loan type. Multiplying by the reported principal value and the assumed duration of the program yields the total value of missed payments that we attribute to the program. For the United States we also partially rely on the estimates in Cherry and others (2021).

European Union. For the EU we primarily rely on information provided by the European Banking Authority (EBA) in their 2020 report to estimate the aggregate reduction in loan payments for households and NFCs in the twenty-five European countries the EBA covers. The headline statistic is that in total about EUR 871 billion of EBA-compliant loans had been granted moratoria as of June 2020, with EUR 860 billion going to households and NFCs.[13] In addition, the EBA reports that other

13. EBA-compliant loans satisfied certain criteria that gave banks regulatory relief (see section III.C).

COVID-19-related relief measures such as noncompliant moratoria and contractual modifications or refinancing applied to an additional EUR 60 billion of loans.

The EBA data further break out the coverage of NFCs and households, with about EUR 495 billion of the loans subject to EBA-compliant moratoria going to NFCs and EUR 365 billion going to households. Overall, 16 percent of SME loans were granted moratoria, followed by 12 percent of commercial real estate loans and 7 percent of residential mortgage loans.

For our estimate of the total payment reductions by households, we assume that all loans are amortizing mortgages, with an original maturity of twenty-four years, an annual interest rate of 2 percent, and monthly payments. Under those assumptions, the monthly mortgage payment is EUR 437 per 100,000 of initial loan principal or 5.3 percent of initial principal annually. Because the loans are amortizing, the mortgage payment per dollar of outstanding principal also depends on the age of the loan. It increases to 6.3 percent of remaining loan principal in five years and 8.2 percent in ten years. Assuming that 25 percent of the loans are fairly new, 25 percent are ten years old, and the rest are five years old implies an average payment as a percentage of principal of 6.5 percent. We further assume that no payments were made for nine months. Taken together, this implies that households had an additional EUR 17.8 billion ($.065 \times 9/12 \times$ EUR 365 billion) of available funds due to these programs.

A similar calculation provides an estimate of the total payment reductions by NFCs. Business loans typically have a much shorter maturity than residential mortgages, ranging from less than a year to ten years or more for some commercial real estate. We assume that all loans are amortizing, with an original average maturity of four years, an annual interest rate of 5 percent, and monthly payments. This implies a monthly payment of EUR 2,303 per 100,000 of initial loan principal or 27.6 percent of principal annually. Because of amortization, the payment amount increases to 52.6 percent of remaining loan principal in two years. We take the 52.6 percent to be the typical payment foregone as a percentage of remaining principal, on an annual basis. We again assume that the payment holiday lasts for nine months. This implies that European NFCs had an additional EUR 195.3 billion ($.526 \times 9/12 \times$ EUR 495 billion) of available funds because of these programs.

Information provided on the largest individual country programs (Italy, France, and Spain) allows those results to be broken out from the European totals. Lacking a breakdown of the share of household versus NFC principal for individual countries, we attribute the share of the total estimated

Table 2. Summary of Imputed Payment Reductions for Four European Countries

France	French banks reported EUR 255 billion of household and NFC loans under moratoria, comprising 7 percent of total loans for households and NFCs. The share of the European total is 255/860 = 29.7 percent, implying a payment reduction of .297 × 213 = EUR 63.2 billion.
Spain	Spanish banks reported EUR 187 billion under moratoria, comprising 10 percent of total loans for households and NFCs. The share of the European total is 187/860 = 21.7 percent, implying a payment reduction of .217 × 213 = EUR 46.3 billion.
Italy	Italian banks reported EUR 156 billion in loans under moratoria, comprising 13 percent of total loans for households and NFCs. The share of the European total is 156/860 = 18.1 percent, implying a payment reduction of .181 × 213 = EUR 38.7 billion.
Germany	While EBA-compliant measures in Germany were less relevant than in other major EU countries, German banks reported the largest amount of loans with other types of COVID-19-related forbearance measures, totaling EUR 14 billion, or 1 percent of total loans. Applying the weighted average payment reduction between household and NFC loans calculated for the EU overall implies missed payments of EUR 3.5 billion (based on .33 × 9/12 × EUR 14 billion).

Source: Authors' calculations.
Note: Estimates based on data and commentary in European Banking Authority (2020).

payment reduction of EUR 213 billion based on a country's share of total loans under moratoria. The calculations and conclusions for Italy, France, Spain, and Germany are summarized in table 2.

Some smaller countries relied the most heavily on moratoria in percentage terms. The EBA notes that the country that had the largest share of bank loans covered by a moratorium was Cyprus at 50 percent, followed by Hungary and Portugal each at greater than 20 percent. The overall average of bank loans to households and NFCs covered under moratoria was 7.5 percent.

Clearly there is considerable uncertainty surrounding these point estimates. The assumed maturity and interest rate on mortgages is loosely consistent with data reported by Statista for recent years for the larger European economies. To the extent that some mortgages had a shorter initial maturity than the assumed twenty-four years or that other types of loans are included with shorter maturities, the estimates are downward biased. The sensitivity to the assumption about rates is reduced by the low and falling mortgage rate environment that has prevailed in Europe for the last decade. The largest source of uncertainty is how long forbearance was in force. According to the EBA, 85 percent of the programs were scheduled to end by year-end 2020 but many were extended into 2021 and some of the arrangements could have continued for even longer. The total payment

reduction is proportional to the length of the forbearance period assumed, making it easy to evaluate the sensitivity to this assumption. The estimates for NFCs have considerable uncertainty associated with the size of the payment relative to loan principal, arising from not having data on the distribution of loan maturities.

United Kingdom. The United Kingdom instituted a payment holiday for residential mortgages in March 2020 that initially allowed a suspension of mortgage payments for up to three months without an effect on borrower credit scores. The policy was soon amended to allow six months of nonpayment. According to Albuquerque and Varadi (2022), at the peak, in May 2020, around 17 percent of all mortgages were on a payment holiday. That proportion declined gradually over time, reaching about 2.5 percent in October 2020. There does not appear to have been large-scale forbearance on business loans.

As for the EU and consistent with information on UK mortgages, the estimated payment reduction is based on an amortizing mortgage with a twenty-four-year maturity bearing an interest rate of 2 percent, with monthly payments. The resulting annual payment per unit of original principal after five years is 6.33 percent of loan principal. The average holiday was assumed to last for four months. Outstanding UK mortgage debt in 2020 totaled about GBP 1.5 trillion. The implied total payment reduction is GBP 31.65 billion.

United States. Under the CARES Act that was enacted on March 27, 2020, US mortgage borrowers could apply for forbearance on loans backed by the government via Fannie Mae, Freddie Mac, or FHA/VA/USDA, and forbearance could last for up to eighteen months inclusive of any extensions. The CARES Act also placed a moratorium on federal student loan payments that is still in effect and that forgives the missed interest payments. Other forbearance measures were instituted as well, but the amounts involved were much smaller.

Cherry and others (2021) report that between March 2020 and May 2021, more than 70 million consumers with loans worth $2.3 trillion entered forbearance, missing $86 billion of their payments.[14] They break the total down into student loans ($45 billion), mortgages ($31 billion), auto debt ($5.7 billion), and revolving debt ($4.7 billion). We take $86 billion as the cumulative total through May 2021.[15]

14. This period is considered to be fourteen rather than fifteen months in our calculations because the CARES Act became law at the end of March.

15. This includes some voluntary forbearance by lenders that shouldn't be included in measures of government support, but we have not adjusted the totals for that effect.

To extrapolate to the end of 2021 for mortgages, we refer to the more recent analysis of Sánchez and Wilkinson (2022) who report that 85 percent of these borrowers had mortgages in forbearance for a year or less. That suggests that less than 15 percent of the mortgages accounted for by Cherry and others (2021) were still in forbearance between June and December 2021. We estimate that an additional $1.6 billion of payments were missed during this time, based on assuming that 10 percent of the mortgages on average were still in the program and prorating it by the relative number of months covered ($31 × .1 × 7/14).

We also need to estimate the additional value of foregone payments for student loans after May 2021, and we do so in two ways. For symmetry with most other pandemic programs, we extrapolate only to December 2021.[16] We employ a back-of-the-envelope calculation similar to what we did for mortgages. Federal student loans outstanding in 2021 stood at approximately $1.5 trillion. About half of loans would have been in grace, deferral, or forbearance even during ordinary times in the past, and the increase in income-based repayment in recent years means that many payments are reduced relative to the originally scheduled amounts. To capture this, we assume that only one-third of the loans would have been in repayment absent the CARES Act. The average interest rate on federal student loans since 2006 has been 6 percent. The standard payment term is ten years but that can be extended to twenty or twenty-five years and extensions happen frequently. Based on these observations, we assume that the typical student loan benefiting from the moratoria has been in repayment for five years and that the principal is originally amortized over twenty years at a 6 percent interest rate. That implies an annual payment equal to 8.6 percent of original principal and 10.1 percent of remaining loan principal after five years. The estimated total missed student loan payments from June 2021 to December 2021 is $29.5 billion ($1.5 trillion/3 × .101 × 7/12). A simple extrapolation from estimates in Cherry and others (2021) serves as a point of comparison and suggests a somewhat smaller total of $22.5 billion ($45 × 7/14) over that period.

Adding the estimated additional $1.5 billion in missed mortgage payments and $29.5 billion in missed student loan payments to the $86 billion in foregone payments reported by Cherry and others (2021) implies total payment reductions of $117 billion.

16. That choice excludes the substantial payment reductions that occurred in 2022 that could be relevant for some questions. We may revisit this in a subsequent draft.

Japan. We have not found mention of any significant loan forbearance policies in Japan. A rental relief program was instituted where participants did not incur any obligation for repayment. There was also tax relief on mortgages. Both of those policies fall into the fiscal category.

ASSESSING COST AND COST INCIDENCE OF FORBEARANCE POLICIES Having previously touched on the potential pros and cons of forbearance policies, here we turn to how to assess their fiscal and private costs conceptually and the incidence of costs. The answer in individual cases will depend critically on how the policy is structured. We consider three major categories of policies: (1) the government forbears on payments owed to itself (e.g., US student loans), (2) the government replaces or provides insurance on payments to private lenders (e.g., the Italian debt moratoria), or (3) the government mandates forbearance by private lenders without compensating them. It is left for future research to apply these principles to quantify the fiscal and private sector costs of the many COVID-19 forbearance policies.

Existing government-backed loans. When a government offers forbearance on its outstanding direct loans or on guaranteed loans that were originated in the past, the incremental fiscal cost of the policy is the difference between the net present value of the associated net government cash flows with and without the forbearance policy (see also sections I.D and II.A). Note that the increase in expected losses following the onset of the pandemic is a sunk cost for the government, and the incremental cost of forbearance does not include those already incurred costs.

A related and more subtle question, both for forbearance policies and for other credit programs, is whether the costs of these policies should be estimated using a static score that abstracts from the macroeconomic effects of the specific policy or whether a dynamic score that takes into account its macroeconomic effects should be used. Importantly, under either approach, cost estimates should take into account the totality of government actions (fiscal, monetary, and regulatory) on the value of the subsidies. Furthermore, a fair value approach to cost estimation incorporates forward-looking expectations about any macroeconomic effects via the market's consensus about the value of risky claims. If government actions are salutary and cause periods of financial stress to be shorter and less severe, incorporating policy effects into projected outcomes will lower the estimated cost, and vice versa for policies that impede recovery. However, given the large uncertainty surrounding the macroeconomic effects of any specific credit program, the two scoring approaches are expected to produce similar cost estimates.

In principle then, the government could institute a forbearance policy with no incremental fiscal cost by adjusting the rates or fees charged so as

to preserve the net present value of future cash flows. A forbearance policy even could have a negative fiscal cost if it caused a net reduction in costly defaults. In practice, however, there are a number of reasons to expect the fiscal cost of such policies to be positive: it extends the period over which the original interest rate subsidies accrue, it increases outstanding loan amounts and thereby the size of potential defaults, it delays recoveries which all else equal reduces their present value, and the policy may be coupled with other costly concessions. A fiscally expensive example is the US moratorium on student loan payments. The program has been extended several times, and the interest component of missed payments is forgiven rather than accrued.

New government backing for private sector loans. The imposition of forbearance policies on private sector lenders and servicers creates costs for those entities that a government can choose to partially or fully absorb. The net present value of the incremental risk absorption by a government is the fiscal cost of the policy. For instance, a government could agree to provide a guarantee to lenders to cover default losses while the moratorium is in effect (e.g., because the borrowing firm goes out of business). The Italian program for bank loans had this feature. It is tricky to estimate the size and incidence of costs with this sort of arrangement. Forbearance tends to reduce near-term defaults but it increases defaults once payments resume. That timing shift makes it relatively inexpensive for the government to guarantee the loans during the forbearance period. That creates additional costs for lenders for the reasons mentioned above (higher balances that default, delayed recoveries, etc.). Those uncompensated costs are equivalent to imposing a tax on lenders, albeit one that is not included in official fiscal accounts.

Mandates on private sector lenders. It appears that under most forbearance policies introduced during the pandemic, lenders received no direct compensation or guarantees for the costs incurred under most mandated forbearance policies. They did receive regulatory relief as described in section II.C. The resulting lender costs can be described as hidden taxes or in budgetary parlance as unfunded mandates. A drawback of uncompensated forbearance mandates is that the costs to lenders are hidden, hard to estimate ex ante, and potentially high. The fact that eventual losses will tend to be larger during longer and deeper downturns is one reason the costs to lenders may be higher than is generally recognized at the onset. Some might argue to the contrary that lender costs are likely to be low or even negative because the policies tend to reduce total losses by providing

time for households and businesses to recover. While that is true for voluntary forbearance, it is reasonable to presume that involuntary policies tend to be costly for lenders.

Finally, while uncompensated forbearance costs are likely to be borne by lenders and their customers most of the time, in the event of a protracted downturn and elevated default rates, some of the costs could again land on governments through their support of distressed financial institutions.

II.C. Regulatory Accommodations

Missed loan payments typically have adverse consequences for both lenders and borrowers even when mutually agreed to. The bank has to report the loan as nonperforming, which affects its capital position, and the borrower's credit score takes a hit. Because of this, it was anticipated that many banks and borrowers would avoid forbearance programs without changes to these rules.

To encourage participation, certain rules and regulations were temporarily relaxed. For example, the EBA (2020) notes that for EU banks, the application of qualifying moratoria did not automatically trigger either a forbearance classification or nonperforming status of the exposure. They did caution banks to continue to monitor credit risk and to classify loans appropriately under the International Financial Reporting Standards.

By design, credit guarantee programs enabled some very risky firms to borrow that otherwise could not have done so. Despite an influx of risky borrowers via these programs, the effect of the large volume of guaranteed loans was to reduce banks' reported risk-weighted assets (RWAs). The EBA (2020) notes that EU banks reported average RWAs to be 18 percent of the exposure value for loans made under public guarantee schemes, whereas the average RWA was 54 percent for banks' loans to NFCs.

The reduction in risk weights was appropriate from a risk measurement perspective because the guaranteed loans were in fact low-risk for the banks. However, a naive reading of the low level of average RWAs could have given some policymakers the false impression that credit risk in the economy was much lower than it actually was.

In the US context, a feature of the student loan moratorium is that missed payments would not be reported to credit agencies and that delinquent loans would be given a fresh start. Haughwout and others (2020) reported this had the effect of increasing average credit scores of student loan borrowers from 647 in March 2020 to 656 in June 2020, primarily initiated by the "curing" of delinquent loans as they entered forbearance.

Table 3. Resources Provided by Credit, Forbearance, and Fiscal Policies as Percentage of 2020 GDP

Country	*Government credit*	*Forbearance*	*COVID-19 fiscal (IMF)*	*Credit + forbearance*	*Credit + forbearance + fiscal*
DEU	1.5	0.1	15.1	1.6	16.7
ESP	10.9	3.6	8.4	14.5	22.9
FRA	6.9	2.4	9.6	9.3	18.9
ITA	10.2	2.0	10.8	12.2	23.1
JPN	5.2	0.0	16.7	5.2	22.0
UK	4.1	1.2	19.3	5.4	24.6
USA	3.9	0.6	21.5	4.5	25.9

Source: Authors' estimates.

Note: For the United States, the Paycheck Protection Program is moved into the credit category and out of the fiscal category where it was recorded in government accounts. Moving it back to fiscal increases that category to 25.3 percent of GDP and reduces the take-up of government credit to 0.08 percent of GDP.

III. Relation of Credit Policies to Macroeconomic Outcomes

In this section we combine the estimates of incremental cash provided by COVID-19 credit policies with incremental fiscal spending for the seven large countries whose credit policies we have studied. We then examine how those policy choices are related to certain macroeconomic outcomes and their differentials across these countries. Specifically, we look at the relation between the incremental resources provided by credit and fiscal policies, both separately and when added together, on GDP growth, savings increases, and inflation.

III.A. Combined Size of Credit, Forbearance, and Fiscal Policies

Estimates of the aggregate incremental resources provided by large, developed countries during the pandemic in the form of government credit support, forbearance, and fiscal policies are summarized in table 3. Credit support is equated to the take-up (drawn principal amounts) of guaranteed and direct loans reported in section II.A. Forbearance is equated to the total payment reductions reported in section II.B. Incremental fiscal support is based on the IMF estimates that are shown in figure 1.

A striking fact revealed by combining estimates of incremental credit and fiscal policies is the relatively small variation across countries in total incremental resources provided, ranging from 16.7 percent of 2020 GDP for Germany to 25.9 percent for the United States, and averaging 22 percent overall. The coefficient of variation across countries for total incremental resources (credit + forbearance + fiscal) is .15, whereas it is .35 for

fiscal spending, .57 for government credit support, and .94 for forbearance. Very similar magnitudes and patterns emerge when quantities are normalized by 2019 GDP, and those results are not reported here.

Also notable is that when credit support and forbearance policies are added to incremental fiscal expenditures, the United States is no longer an outlier from the rest of the pack. When only fiscal policy is considered, and when the PPP is treated as fiscal, the United States has the highest ratio of fiscal expenditures to GDP, exceeding the second-place United Kingdom by 6 percentage points of GDP. This suggests that narratives emphasizing the role of aggressive US policy in fueling domestic and international inflation may exaggerate the disproportionate effect of US fiscal actions.

It appears then that these countries used credit and fiscal policies as substitutes. This observation raises a number of important questions. Why did certain countries, in particular Spain, France, and Italy, rely so much less on traditional fiscal measures? Was credit more attractive due to more binding constraints on government spending and the lack of transparency about the cost of credit policies? Or was credit support perceived as truly more cost-effective or efficacious? Did the countries that relied more heavily on credit policies get the same growth benefits as those that relied more on traditional fiscal measures? These questions are also addressed in the literature on the effects of fiscal space (Romer and Romer 2019; Bergant and Forbes 2023).[17] While we won't speculate further about the various motivations for countries adopting different policy mixes, in the next section we consider how those choices correlated with differential economic growth, savings rates, and inflation.

III.B. Do Credit Policies Help Explain Macroeconomic Outcomes?

To explore whether taking credit support and forbearance into account helps to explain macroeconomic outcomes, we consider the relation between the differentials in macroeconomic country outcomes and alternative measures of the size of government policy interventions. Table 4 reports real GDP growth, the gross increase in private savings, and inflation by country. For the different variables, the time horizon is chosen as follows: real GDP growth is measured from the fourth quarter of 2020 to the fourth quarter

17. Romer and Romer (2019) find that pre-pandemic countries with more fiscal space responded to downturns more aggressively. Bergant and Forbes (2023) find that during the pandemic fiscal space had only a weak correlation with the announced size of fiscal response. These studies include additional macroeconomic policies and cover a wider set of countries, but they do not explicitly incorporate estimates of the incremental resources provided by credit and forbearance policies.

Table 4. Real GDP Growth and Inflation Rates

Country	*Inflation*	*Real GDP growth*	*Increased private saving*
DEU	0.115	0.012	0.195
ESP	0.073	0.067	0.201
FRA	0.071	0.052	0.109
ITA	0.125	0.065	0.352
JPN	0.038	0.006	0.173
UK	0.111	0.088	0.544
USA	0.078	0.057	0.332

Source: IMF World Economic Outlook Database.

Notes: Inflation is the change in CPI between October 2021 and October 2022. Real GDP growth is cumulative growth of quarterly compounded rates through 2021:Q4 divided by cumulative growth of quarterly compounded rates through 2020:Q4. Increased saving is the difference between gross private saving in 2020 and in 2019, divided by 2020 GDP.

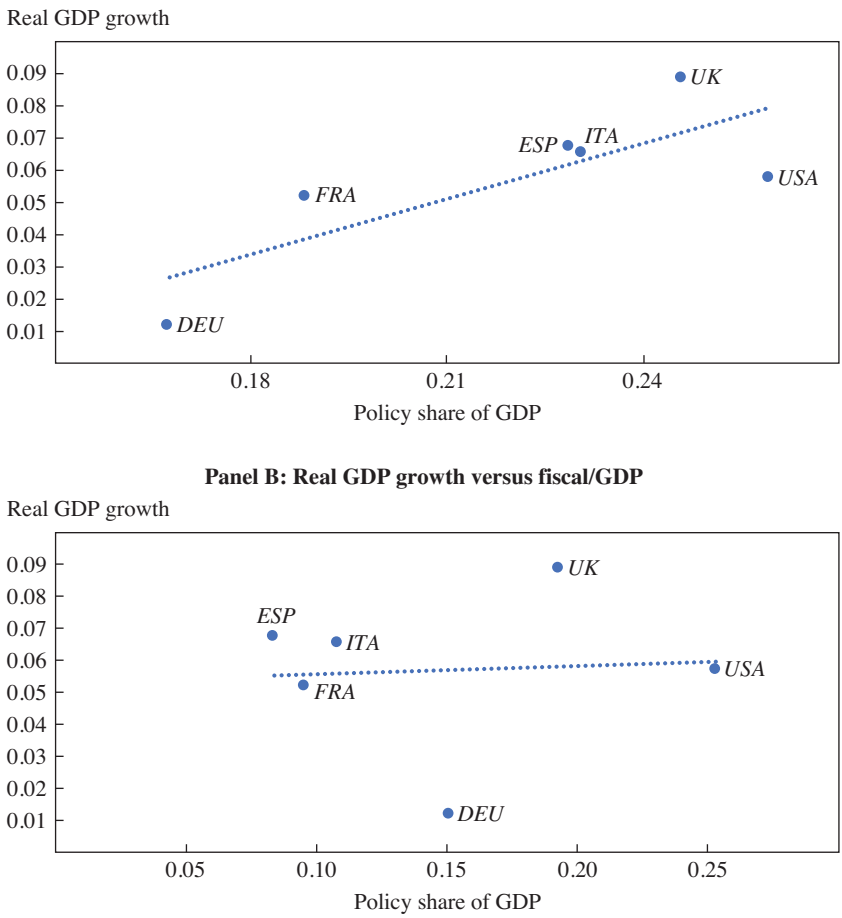
of 2021, a period where pandemic lockdowns were in effect in many countries. Inflation is measured as the ratio of the Consumer Price Index (CPI) in October 2022 to the CPI in October 2021, the period when inflation accelerated. Increased savings is measured as the difference between gross private savings in 2020 and in 2019, divided by 2020 GDP. The reported results are fairly insensitive to modest shifts in the period start date. Because Japan is an outlier in the relation between its policies and macroeconomic outcomes, we exclude it from the figures.

REAL GDP GROWTH The cross-country relation between real GDP growth and the different measures of policy can be summarized by the correlations between them.[18] The correlation between real GDP growth and the sum of the three policy types is .62, whereas it is .35 for the sum of credit and fiscal policy, and −.07 for fiscal policy alone. The unintuitive negative sign for fiscal policy appears to be driven by Japan being an outlier with its very low growth and aggressive fiscal policy.

When the correlations are recalculated excluding Japan, the correlation of real GDP and the sum of the three policy types increases to .79, whereas for credit plus fiscal it is .61, and for fiscal alone it is .07. Figure 4 illustrates the much stronger relation between cross-sectional real growth and a broad measure of resources made available versus the much weaker relation with differences in traditional fiscal spending.

18. We report correlations because with only a few countries and one historical episode, there are insufficient data to undertake a more formal statistical analysis. There are not enough degrees of freedom for even a simple multivariate analysis.

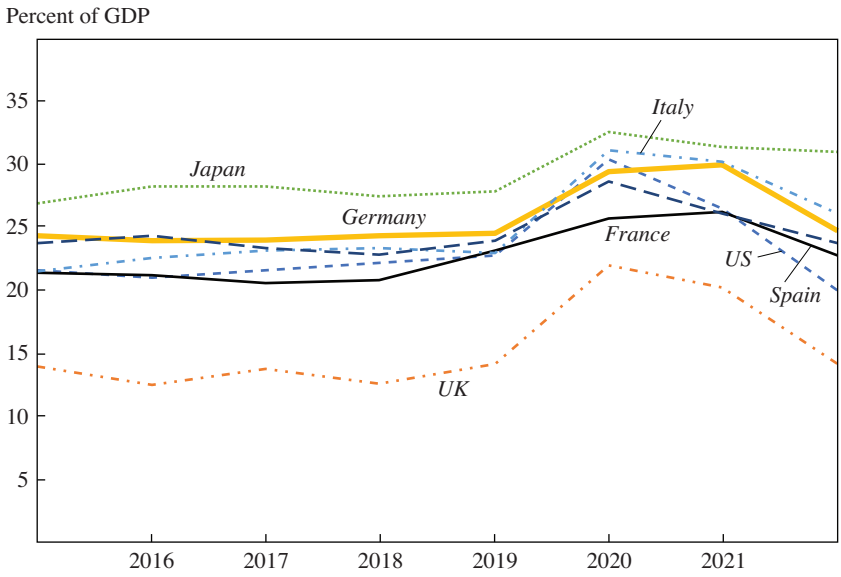
Figure 4. Real GDP Growth versus Alternative Policy Measures



Source: Authors' calculations using the IMF World Economic Outlook Database and IMF Database of Country Fiscal Measures in Response to the COVID-19 Pandemic.

Note: Real GDP growth is cumulative growth of quarterly compounded rates through 2021:Q4 divided by cumulative growth of quarterly compounded rates through 2020:Q4. Fiscal is incremental spending for COVID as reported by the IMF. Credit is take-up in COVID guaranteed and direct lending programs. Forbearance is foregone payments in those programs. All are normalized by 2020 GDP.

Figure 5. Gross Private Savings from 2019 to 2020



Source: IMF World Economic Outlook Database.

GROWTH IN PRIVATE SAVINGS RATES It is well known that savings rates by both firms and households increased sharply during the pandemic. Explanations for elevated savings rates include the reduced activities that money could be spent on, an increased precautionary demand for savings, and the outpouring of funds made available through government spending and credit policies that are the focus here. Figure 5 shows the significant increase in gross private savings that occurred between 2019 and 2020. In most countries savings rates remained elevated in 2021 as well, and only recently have they returned to pre-pandemic rates in most countries. The results reported below on the relation between savings and different policy measures don't qualitatively change when the savings increase is measured as the average of 2020 and 2021 savings relative to 2019.

As for real GDP growth, we consider the correlation between the different measures of policy and increased saving. The correlations are based on the values in tables 3 and 4. The correlation with the sum of the three policy types is .67, whereas it is .69 for the sum of credit and fiscal policy, and .54 for fiscal policy alone. The correlation with forbearance policy alone is −.12. Excluding Japan, the correlation with the sum of the three

policy types increases to .69, whereas for credit plus fiscal it is .79, and for fiscal alone it is .63.

The correlations between savings and policy actions are high (figure 6). Adding take-up of credit support programs increases the correlation, but adding missed payments from forbearance decreases it. A possible explanation for the negative correlation with forbearance is that the more aggressive forbearance policies were a reaction to more-elevated spending needs, and in fact those policies were relatively efficient at getting funds to businesses and households that needed to spend the money. The high correlation with take-up of credit and traditional fiscal policy supports the conjecture that the multipliers on both types of policies were probably lower than in previous recessions.

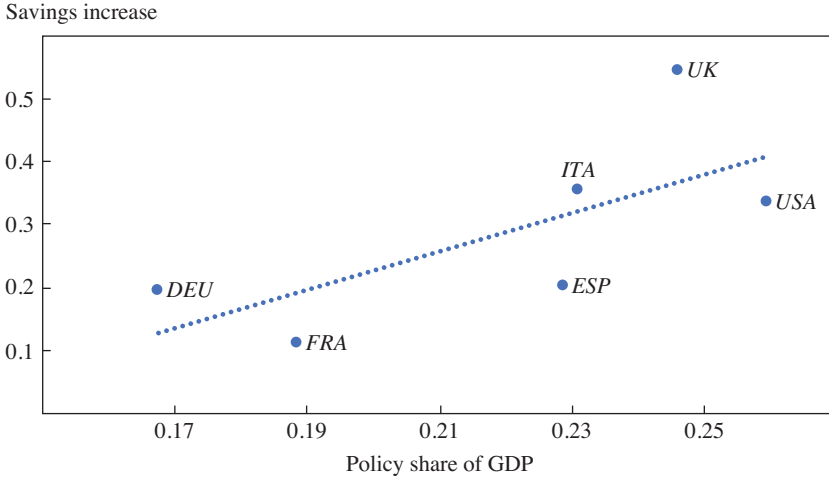
INFLATION The correlation between inflation differentials across countries and any of our policy measures is much weaker than for real GDP growth or for savings rates. It is also lower when credit support and forbearance are included (−.1) than for fiscal policy alone (.17). Japan is excluded in these calculations because Japan's very low inflation juxtaposed with its high fiscal spending causes the correlation between inflation and any of the policy measures to turn sharply negative. A possible reason for inflation differential between these countries to be small is that four of them share a common monetary authority. Figure 7 illustrates the relation between inflation and the different policy measures.

Koch and Doureldin (2023) calculate a measure of unexpected inflation, which is the difference between realized inflation and a projection of inflation imputed from the IMF World Economic Outlook database. Unexpected inflation in 2021 (relative to the October 2019 forecast) is positively correlated with the policy share of GDP, either measured by fiscal policy alone or also including credit support and forbearance.

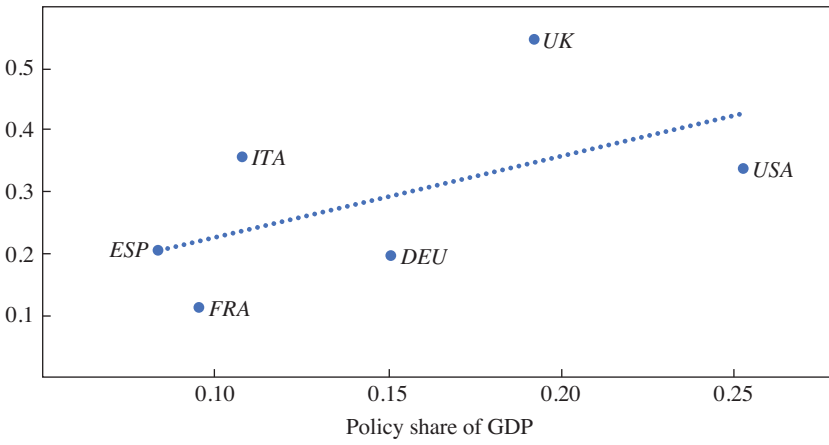
Despite the finding that the take-up rate of credit support and forbearance does not by itself help to explain the cross section of inflation rates across these countries, consideration of the additional resources made available may strengthen the case for government policies having been an important factor in the elevated levels of inflation around the world. The combined measure of incremental government resources averages 22 percent of GDP across the seven countries, versus an average of 14.5 percent for fiscal policy alone. To the extent that credit and forbearance policies contributed to higher savings rates, they could also have caused inflation to continue for longer as those higher levels of savings are gradually run down. The correlation between savings and real GDP growth, and between savings and inflation, are high. The correlations are .65 and .51 respectively when Japan is excluded, and .67 and .56 with Japan. A speculative interpretation

Figure 6. Savings Increase in 2020 versus Alternative Policy Measures

Panel A: 2020 savings increase versus (fiscal + credit + forbearance)/GDP



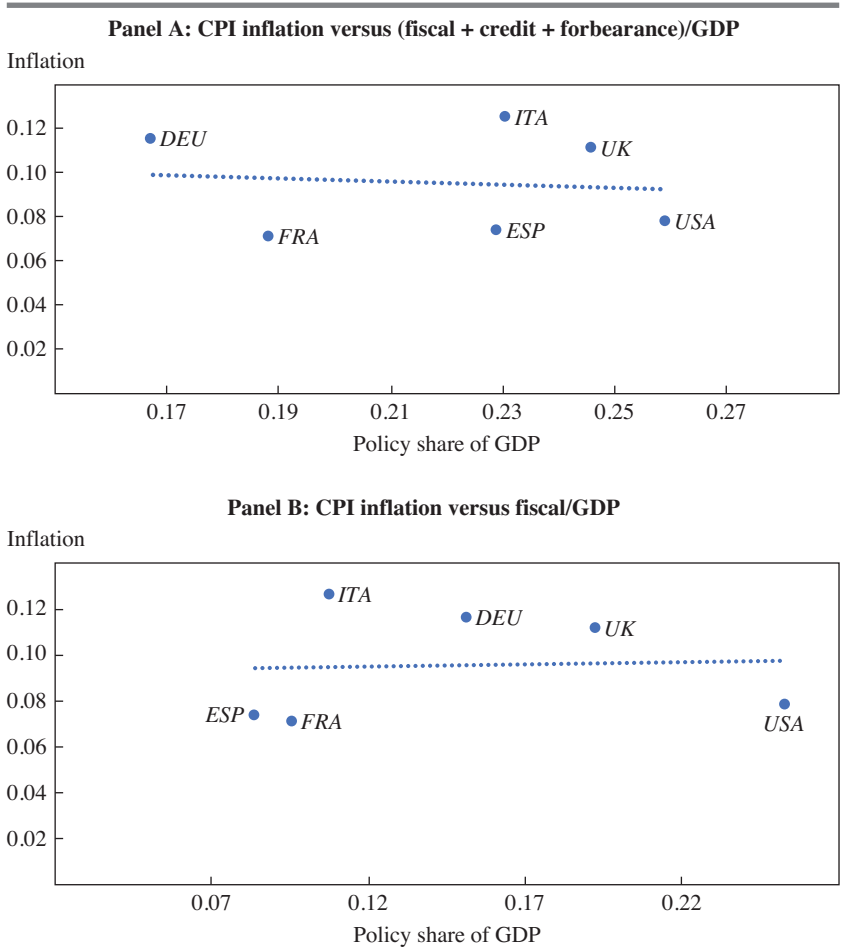
Panel B: 2020 savings increase versus fiscal/GDP



Source: Authors' calculations using the IMF World Economic Outlook Database and IMF Database of Country Fiscal Measures in Response to the COVID-19 Pandemic.

Note: Increased saving is the difference between gross private saving in 2020 and in 2019, divided by 2020 GDP. Fiscal is incremental spending for COVID-19 as reported by the IMF. Credit is take-up in COVID-19 guaranteed and direct lending programs. Forbearance is foregone payments in those programs. All are normalized by 2020 GDP.

Figure 7. Inflation versus Alternative Policy Measures



Source: Authors' calculations using the IMF World Economic Outlook Database and IMF Database of Country Fiscal Measures in Response to the COVID-19 Pandemic.

Note: Inflation is the change in CPI between October 2021 and October 2022. Fiscal is incremental spending for COVID-19 as reported by the IMF. Credit is take-up in COVID-19 guaranteed and direct lending programs. Forbearance is foregone payments in those programs. All are normalized by 2020 GDP.

is that countries that recovered more rapidly wound up saving more of the incremental resources provided by government policies, and that as those higher savings levels are being spent down, they are contributing to higher inflation.

IV. Concluding Remarks

We began by noting the unprecedented number and scale of new credit support and forbearance policies introduced in response to the economic upheaval caused by the COVID-19 pandemic, but also that the effective size of these policies and their macroeconomic effects have not been quantified. The main goal of this project was to begin to fill those gaps in two ways. The first was by developing and implementing procedures to size the different programs in ways that are meaningful for macroeconomic analysis. The second was to use those statistics to look for evidence about whether in fact these policies had quantitatively significant economic effects during the pandemic.

We find that the amount of funds injected into the economy via credit support and forbearance policies was significant: incremental traditional fiscal spending averaged 14.5 percent of GDP across countries, whereas the combined effect of these policies and incremental traditional fiscal spending averaged 22 percent. Furthermore, there was much more uniformity across countries in the combined GDP share of policies than in their individual components. Taking into account the combined size of credit and fiscal policies also appears to better explain cross-country differences in real GDP growth and savings rates than does fiscal policy alone, with the caveat that any inferences are only suggestive because of the small sample size.

The quantitative analysis is restricted to large, advanced economies because of the very limited availability of information on credit support and forbearance programs for most countries. However, the countries included accounted for 45 percent of global GDP in 2020 and an even larger share of global government fiscal and credit market interventions, suggesting that the policies studied here were of quantitative significance globally. We hope that future researchers will undertake similar analyses for additional countries.

The conceptual discussions regarding the measurement of fiscal costs and stimulus effects, the transmission channels through which credit policies operate, and of the pros and cons of fiscal versus different types of credit policies pertain to EMDEs as well as to advanced economies. However, some of the negative aspects of credit policies that we alluded to but did not

emphasize may be particularly salient for EMDEs. That includes their lack of transparency in government accounting for credit and the hidden fiscal risks that are likely to be realized when the economy is weak and fiscal resources are scarce. EMDEs may have more-severe transparency problems, less ability to evaluate the risks, and are likely to have less capacity than advanced economies to manage such shocks when they materialize.

Economists generally do not give credit policy the standalone status accorded to monetary and fiscal policies. An unfortunate side effect of that omission is that the costs and other information about credit policies are poorly and inconsistently measured and reported on in official statistics, and sometimes omitted altogether.[19] There is also less of a common understanding of the ways in which credit policies, as distinct from conventional monetary and fiscal policies, affect macroeconomic and future fiscal outcomes. As we have emphasized in this paper, all of this obscures the effects arising from the totality of credit policies, including the significant fiscal costs and risks they entail. This paper, we hope, is a start at demonstrating the feasibility and the benefits of taking a more holistic approach to credit policies.

ACKNOWLEDGMENTS We thank Alan Auerbach, Janice Eberly, and Carmen Reinhart for their insightful comments and suggestions on an earlier draft. We also thank the participants at the *BPEA* Spring 2023 Conference. Any opinions expressed are those of the authors and do not necessarily reflect those of the IMF, its executive board, or its management.

19. These issues are discussed in more detail in Hong and Lucas (2023) and references therein.

References

Albuquerque, Bruno, and Alexandra Varadi. 2022. "Consumption Effects of Mortgage Payment Holidays during the Covid-19 Pandemic." Working Paper 963. London: Bank of England. https://www.bankofengland.co.uk/working-paper/2022/consumption-effects-of-mortgage-payment-holidays-evidence-during-the-covid-19-pandemic.

Anderson, Julia, Francesco Papadia, and Nicolas Véron. 2021. "COVID-19 Credit Support Programs in Europe's Five Largest Economies." Working Paper 21-6. Washington: Peterson Institute for International Economics. https://www.piie.com/publications/working-papers/covid-19-credit-support-programs-europes-five-largest-economies.

Auerbach, Alan J., Yuriy Gorodnichenko, Peter McCrory, and Daniel Murphy. 2021. "Fiscal Multipliers in the COVID19 Recession." Working Paper 29531. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29531.

Bergant, Katharina, and Kristin Forbes. 2023. "Policy Packages and Policy Space: Lessons from COVID-19." Working Paper 31254. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w31254.

British Business Bank. 2021. "Record Levels of Smaller Businesses Sought External Financial Support in 2020, with Further Significant Demand Expected in 2021, Finds British Business Bank Research." Press release, March 10. https://www.british-business-bank.co.uk/press-release/record-levels-of-smaller-businesses-sought-external-financial-support-in-2020-with-further-significant-demand-expected-in-2021-finds-british-business-bank-research/.

Cherry, Susan, Erica Jiang, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2021. "Government and Private Household Debt Relief during COVID-19." *Brookings Papers on Economic Activity*, Fall, 141–199.

Core, Fabrizio, and Filippo De Marco. 2021. "Public Guarantees for Small Businesses in Italy during COVID-19." Discussion Paper DP15799. London: Centre for Economic Policy Research. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3784039.

European Banking Authority (EBA). 2020. *First Evidence on the Use of Moratoria and Public Guarantees in the EU Banking Sector*. Paris: EBA. https://www.eba.europa.eu/sites/default/documents/files/document_library/Risk%20Analysis%20and%20Data/Risk%20Assessment%20Reports/2020/Thematic%20notes/Thematic%20note%20on%20moratoria%20and%20public%20guarantees/936761/For%20publication%20-%20Thematic%20note%20on%20moratoria%20and%20public%20guarantees.pdf.

Haughwout, Andrew F., Donghoon Lee, Joelle Scally, and Wilbert van der Klaauw. 2020. "A Monthly Peek into Americans' Credit during the COVID-19 Pandemic." Liberty Street Economics, Federal Reserve Bank of New York, August 6. https://libertystreeteconomics.newyorkfed.org/2020/08/a-monthly-peek-into-americans-credit-during-the-covid-19-pandemic.html.

Hong, Gee Hee, and Deborah Lucas. 2023. "Evaluating the Costs of Government Credit Support Programs during COVID-19: International Evidence." Working

Paper 2023/16. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2023/01/27/Evaluating-the-Costs-of-Government-Credit-Support-Programs-during-COVID-19-International-528660.

Koch, Christoffer, and Diaa Noureldin. 2023. "How We Missed the Inflation Surge: An Anatomy of Post-2020 Inflation Forecast Errors." Working Paper 2023/102. Washington: International Monetary Fund.

Lucas, Deborah. 2016. "Credit Policy as Fiscal Policy." *Brookings Papers on Economic Activity*, Spring, 1–41.

Lucas, Deborah. 2020. "Policy Rx for the Economy: Cash or Credit?" In *Impact of COVID-19 on Asian Economies and Policy Responses*, edited by Sumit Agrawal, Zhiguo He, and Bernard Yeung. Singapore: World Scientific.

Remy, Mitchell, Deborah Lucas, and Damien Moore. 2011. "An Evaluation of Large-Scale Mortgage Refinancing Programs." Working Paper 2011-4. Washington: Congressional Budget Office. https://www.cbo.gov/sites/default/files/112th-congress-2011-2012/workingpaper/09-07-2011-Large-Scale_Refinancing_Program_1.pdf.

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.

Sánchez, Juan M., and Olivia Wilkinson. 2022. "Forbearance during COVID-19: How Many Borrowers Used It, and for How Long?" Blog post, May 31, On the Economy, Federal Reserve Bank of St. Louis. https://www.stlouisfed.org/on-the-economy/2022/may/forbearance-covid19-how-many-borrowers-how-long.

Comment and Discussion

COMMENT BY

ALAN J. AUERBACH The COVID-19 pandemic has left us with myriad terrible outcomes, but for the economist interested in studying government policies, it has also provided a wealth of new data. Going beyond massive fiscal interventions and accommodating monetary policies, governments around the world introduced a wide variety of credit programs aimed at sustaining economic activity and promoting individual well-being. While there have been attempts to collect information about these credit programs, notably by the International Monetary Fund (IMF), as cited in the paper, Hong and Lucas take an important step beyond this in quantifying the effective size of these credit programs as well as their potential impacts on government budgets and economic activity. In doing so, they make the case for paying more attention to credit programs and their effects, even in more normal economic times, and for treating credit programs separately from other types of government policies.

WHAT MAKES CREDIT POLICIES DIFFERENT? Why should we consider credit policies separately? Hong and Lucas offer several justifications. First, credit policies have different effects on the budget than traditional fiscal policies. Whereas the budgetary cost of a tax cut or an increase in government spending equals the upfront change in tax revenues or purchases, the authors argue, the budgetary cost of a credit program is typically much smaller. For example, a government loan program wouldn't incur a cost

Editors' Note: Carmen Reinhart provided a thoughtful discussion on the conference version of the paper by Hong and Lucas in this volume at the Spring 2023 *BPEA* Conference. The recording of her discussion is available at https://www.brookings.edu/events/bpea-spring-2023-conference/.

Brookings Papers on Economic Activity, Spring 2023: 334–345 © 2024 The Brookings Institution.

equal to the amount of initial lending, but just the net present value cost of the loans made once account is taken of future payments of interest and principal. For most of the credit programs evaluated here, this net cost is only a fraction of funds initially made available. This opens the question of how one should measure the net future payments to the government, and Hong and Lucas forcefully argue for using a "fair value" basis for such measurement, relying on risk adjustment rather than simply discounting future payments at the safe government interest rate. While the fair value approach tends to increase credit programs' budgetary costs, there remains a large gap between budgetary cost and initial outlay of funds.

But is this comparison with traditional fiscal policy accurate? That is, should one think of changes in government spending and taxes as happening independently over time, so that each policy innovation occurs only in one year, with no complementary or offsetting components in the future? For many fiscal policies, this is not the case. An example that is perhaps closest to a credit policy in its pattern over time is the change in income tax withholding implemented by President George H. W. Bush in 1992, which was offset when taxpayers filed their returns for that year in 1993—essentially an interest-free loan to taxpayers.[1] While one might dismiss this example as being fairly unusual, in that it was implemented by executive order and offset within roughly a year, it is hardly unique. The use of accelerated depreciation deductions for business investment—going back over two decades in its most recent form, bonus depreciation—involves giving businesses tax deductions earlier than they otherwise would have been permitted, meaning that the time pattern of deductions is accelerated but the sum of nominal deductions over time is unaffected—again, an interest-free loan by the government, although in this case one that takes many years to unwind and has been implemented through the legislative process.

Indeed, we may think of current fiscal policy shocks as having future elements, even in cases where the legislation giving rise to the policy change does not explicitly include such future components. This is the basic approach followed when we estimate the effects of a government policy shock in a structural vector autoregression (SVAR) model, in which we trace out the effect of the shock not only on, say, GDP but also on the future values of the policy variable itself and measure the effects of the policy shock on output considering not only the initial change in the policy variable but also those future induced policy changes. In this context, we

1. Survey evidence in Shapiro and Slemrod (1995) suggests that this temporary policy did lead to an increase in consumption among a substantial fraction of the population.

are defining the dynamic shape of policy using a reduced form approach, rather than relying on what is explicit in legislation—if a current reduction in taxes leads to future increases in taxes, for example, we treat this as a single policy change. Indeed, one could take the same perspective for credit programs. Suppose, for example, that a temporary holiday on student loan repayments makes future reductions in student loan payments more likely through additional loan relief or the restructuring of loans. Should one ignore the future consequences of current policy actions?

In summary, it is not clear to me that one should view credit policies as having a smaller budget footprint because they involve future offsetting payments to the government. This may make them more cost effective than some traditional fiscal policies, but the difference in character is easy to overstate.

A second difference between credit policies and traditional fiscal policies that Hong and Lucas discuss is that credit policies can be effective at relaxing temporary borrowing constraints. The intuition here is straightforward. If the targeted group of businesses or individuals faces borrowing constraints, a policy that makes loans available, either through direct government provision or through government guarantees or subsidies in the private loan market, can expand access to credit. Here, again, one should recognize that certain traditional fiscal policies may share the characteristic with credit policies. Consider once more the case of accelerated depreciation deductions, which in many studies over time have been found to increase investment. The increase in investment could be induced through more than one channel; not only does the effective interest-free loan lower the user cost of capital, the standard way in which investment incentives are thought to encourage investment, but it also makes funds available immediately, especially in the extreme case of full investment expensing, where the entire deduction is received in the year of investment. Indeed, Zwick and Mahon (2017) argue that it is the latter effect that drives the largest responses to investment incentives, which suggests an impact through the relaxation of borrowing constraints.[2]

Not only do some traditional fiscal policies serve to relax borrowing constraints, some credit policies do not seem particularly formulated with that objective in mind. Hong and Lucas are critical of the US student loan

2. An attribute of accelerated depreciation policies that works against its effectiveness at relaxing borrowing constraints is the limit on deductibility for companies without adequate taxable income, which can be particularly acute during periods of economic weakness; see, for example, Altshuler and others (2009). However, this limit is also a policy choice, not a necessary attribute of investment incentive policy.

relief adopted during the pandemic, arguing that the relief plan was poorly targeted to relieve pressure on those in need.

Finally, credit programs are treated differently in government budgets. This difference was the motivation behind the Federal Credit Reform Act of 1990, which required that subsidies embedded in federal credit programs be accounted for on an accrual basis. As a consequence of this legislation, major credit legislation involving anticipated subsidies has immediate budgetary consequences. Thus, even if the costs of such programs are realized after the end of the budget window, these costs are accounted for. Hong and Lucas argue that the current US accounting standard for federal credit programs is too lax, as it fails to adjust for risk. But, even with its current policy, the United States is an outlier. According to the authors, most other countries do not use accrual accounting for their credit programs, recognizing their budgetary costs only when they are realized, that is, on a cash flow basis. Moreover, as these programs are typically treated as below-the-line activities outside the standard scope of government budgeting, there may not be any immediate budgetary account taken, even of those costs that will eventually show up in the future. From a political economy perspective, this is a potentially important difference between credit policies and traditional fiscal policies. Governments that face formal budget restrictions, or that may want to hide the fiscal costs of their policy actions, could move to adopt credit policies not because these policies have better design attributes for the particular challenges to be addressed but because of accounting differences. It would seem that reforms of budgetary accounting, rather than just tinkering with fiscal rules themselves, should be on the agenda, especially outside the United States.

A different type of misleading accounting shows up in one particular type of credit program that Hong and Lucas evaluate—forbearance policies that involve putting off or forgiving periodic payments that otherwise would be due, including temporary relief from paying rent or making payments on mortgages or student loans. A key issue here is who absorbs the costs of the forbearance. According to the authors, many forbearance programs adopted during the pandemic imposed all or most of the costs on the private sector, as creditors and landlords did not receive payments that were due and governments offered little or no compensation for the missed payments. These policies also might have been attractive to governments seeking to understate the true costs of programs, for they involve hidden taxes on one group of private agents to fund transfer payments to another group of private agents. Again, there might be other reasons why such policies might make sense. For example, from an ex ante insurance

perspective, those implicitly being taxed may be in a better position to bear such taxes than other groups when events like the pandemic occur. But in the current setting, it is hard to know the relative importance of such factors and budgetary treatment in decisions to adopt forbearance policies.

WHAT THE PAPER FINDS Starting from a very useful data set compiled by the IMF measuring the magnitudes of above-the-line (fiscal) and below-the-line (credit) policies adopted by a large number of countries from the start of the pandemic approximately through the third quarter of 2021, Hong and Lucas adjust the credit policy measures for seven large countries that accounted for the lion's share of credit policies during the period: the United States, Germany, Japan, the United Kingdom, Italy, France, and Spain, making additions and subtractions based on a very detailed study of the credit programs adopted in these countries. The description of these adjustments takes up a significant part of the paper, and they represent an enormous amount of work on the part of the authors that result in very large adjustments relative to the numbers drawn from the database. For the two main types of credit policies—(1) credit guarantees and direct lending programs, and (2) forbearance policies—they estimate the direct impact (the value of loans taken up in the first case; the value of delayed or foregone payments in the second). For programs in the first category, they also provide measures of fiscal costs, following their preferred fair value approach.[3]

This deep dive into the seven countries' pandemic programs yields a number of interesting findings. First, credit program take-up was substantially smaller than the envelope size, that is, funds potentially available. This is especially true of Germany, for which committed funds for credit programs, at 1.5 percent of GDP, are less than one-tenth of the envelope. Second, the estimated fiscal cost was 37 percent of the total funds actually committed under these credit programs, that is, nearly two-thirds of the funds initially extended by the government were recouped. Third, even with the adjustment of program size by considering funds actually committed, rather than program envelopes, government credit programs were much larger in magnitude than forbearance policies, the latter ranging from zero to 3.6 percent of GDP compared to a range of 1.5 percent to 10.2 percent of GDP for the former. Fourth, credit programs seem to be substitutes for regular fiscal policies, in the mechanical sense that countries that did more of the first did less of the second—there is much less variation among countries in the sum of the two than in either separately. But we don't really

3. They also discuss regulatory accommodations adopted during the pandemic but do not provide monetary estimates of the relief provided by such actions.

know why some countries relied more heavily on traditional fiscal policies and others on credit policies, or why the take-up of credit policies was so low in Germany and so high (relative to the envelope) in others, notably Spain. Did government borrowing restrictions play a role in restricting the use of government credit? Did budget rules push some countries away from traditional fiscal policies? Was the take-up of credit policies higher in countries where subsidies (i.e., fiscal cost as a fraction of committed funds) were higher? There is some speculation in the paper on these issues, but a more systematic analysis would be a useful topic for further research.

In considering these questions, it is interesting to look at the larger group of countries in the original IMF data. While such data do not have the adjustments made by Hong and Lucas for the size of credit programs, they may still be useful for discerning patterns of behavior. Notably, advanced economies, in the aggregate, not only engaged in more of both credit programs and traditional fiscal programs than emerging market economies (between 11 and 12 percent of GDP for each type of program for advanced economies, and less than half that in both categories for emerging market economies), but it also appears that emerging market economies relied relatively less heavily on credit programs.[4] Could the strength of a country's financial infrastructure be playing a role, even among the seven economies on which this paper focuses, in determining reliance on credit market programs?

After discussing the sizes of credit programs across countries, Hong and Lucas move on to what might be the most important questions for evaluating the programs: What were their macroeconomic effects, and how did these effects differ for different types of credit programs? Unfortunately, with only seven countries in the sample, the conclusions one can draw are limited, which leaves the authors in a challenging situation. They opt for the compromise of presenting simple correlations accompanied by caveats to the reader about making causal inferences. Even in this limited context, there are some interesting findings to contemplate. First, there is a strong correlation between GDP growth during 2021 and the combined size of fiscal and credit programs. Second, there is a low correlation between GDP growth and fiscal policy size alone. Third, there is a strong correlation between increases in private saving between 2019 and 2020 and both fiscal

4. For advanced economies as a group, the percentages of GDP are 11.7 for fiscal policies and 11.4 for credit policies. For emerging market economies, the percentages were 5.7 and 4.2, respectively. Data from International Monetary Fund, "Database of Fiscal Policy Responses to COVID-19," https://www.imf.org/en/Topics/imf-and-covid19/Fiscal-Policies-Database-in-Response-to-COVID-19.

policy size and combined fiscal and credit policy size. Finally, there is essentially no relationship between policy size and Consumer Price Index inflation between October 2020 and October 2021.

Hong and Lucas are not entirely successful in heeding their own admonitions about drawing causal inferences from these thought-provoking results, and neither likely will be very many readers. Therefore, it is useful to think about what might lead one astray when drawing what might seem to be obvious conclusions from these results. A first problem is endogeneity. Why did the size of policy interventions vary across countries? If traditional fiscal policies were adopted as a countercyclical tool, for example, then countries that suffered larger initial shocks to output at the outset of the pandemic might have adopted more expansive fiscal policies, pushing toward a negative correlation between output growth and fiscal policy size. Delaying the timing of the measurement of output growth, as the authors do (looking only at growth starting in the last quarter of 2020, after a substantial share of the fiscal policies had already been adopted), might help avoid some of this spurious correlation.[5] But there isn't a real solution without some independent source of policy variation—an instrument for policy actions undertaken during the pandemic that is unrelated to the severity of the pandemic-induced economic downturn.[6]

A second problem in interpreting these results is that the small sample, though including many large countries, may be unrepresentative for the larger set of all advanced economies. For example, in the original IMF data set, the correlation between the 2021 year-on-year GDP growth rate among all advanced economies and the magnitude of pandemic fiscal interventions is +0.54, whereas it is –0.08 for the seven countries in the sample used by Hong and Lucas.[7]

Third, even if the correlation coefficient accurately measures a causal relationship, it lacks the scaling that one would want to measure the impact of one variable, in this case the combined size of fiscal and credit policies, on the other—GDP growth; a regression coefficient, rather than a correlation coefficient, serves this purpose. For the seven-country sample,

5. Indeed, based on the authors' data, the correlation between output growth and fiscal policy size is more negative if one measures the growth rate through the fourth quarter of 2021 starting in the second quarter of 2020, rather than the fourth.

6. One such instrument might be a country's bond rating, which Romer (2021) found to be predictive of the size of a country's fiscal interventions during the pandemic.

7. The 2021 year-on-year growth rate is used here because the publicly available IMF data set has GDP data only at an annual frequency. The advanced economy sample excludes Taiwan, for which there is no information in the pandemic policy data set.

the coefficient in a regression of GDP growth on the sum of fiscal and credit policy sizes is 0.57 (with a standard error of 0.33), which is a multiplier of reasonable size, given the literature, particularly in light of the fact that effects on output likely were smaller as a result of restricted movement during the pandemic (Auerbach and others 2022).

In short, there are a lot of intriguing findings about the potential effectiveness of credit policies during the pandemic, but it would require a larger sample of countries and a more careful empirical methodology to draw any serious conclusions. It is hardly the fault of Hong and Lucas that we don't have adjusted data for more countries on credit policies, given the effort they have expended to produce what they have, but one lasting contribution of their work could be to spur others to implement their methodology to produce larger data sets for analysis.

REFERENCES FOR THE AUERBACH COMMENT

Altshuler, Rosanne, Alan J. Auerbach, Michael Cooper, and Matthew Knittel. 2009. "Understanding U.S. Corporate Tax Losses." *Tax Policy and the Economy* 23, no. 1: 73–122.

Auerbach, Alan J., Yuriy Gorodnichenko, Peter B. McCrory, and Daniel Murphy. 2022. "Fiscal Multipliers in the COVID19 Recession." *Journal of International Money and Finance* 126:102669.

Romer, Christina D. 2021. "The Fiscal Policy Response to the Pandemic." *Brookings Papers on Economic Activity*, Spring, 89–110.

Shapiro, Matthew D., and Joel Slemrod. 1995. "Consumer Response to the Timing of Income: Evidence from a Change in Tax Withholding." *American Economic Review* 85, no. 1: 274–83.

Zwick, Eric, and James Mahon. 2017. "Tax Policy and Heterogeneous Investment Behavior." *American Economic Review* 107, no. 1: 217–48.

GENERAL DISCUSSION Robert Hall suggested that the economics of insurance might help quantify the effects of credit programs. Hall noted that the basic principle of insurance is the stabilization of marginal utility. He argued that the United States government's credit programs succeeded in stabilizing marginal utility by stabilizing consumption and that some parts of the credit programs also closely resembled insurance because of their repayment requirements.[1] He ended his remarks by suggesting that

1. Federal Reserve Bank of St. Louis, "Personal Consumption Expenditures: Total for United States," https://fred.stlouisfed.org/series/USPCE.

one could quantify the benefits of pure credit programs by measuring the change in marginal utility, which was pushed nearly to zero by the programs when it otherwise would have been large.

Amy Finkelstein also commented on the potential insurance value of the credit programs, noting that there are two possible explanations for the variable take-up rates of the programs across countries, most of which were well below 100 percent. First, frictions such as transaction costs and lack of information prevented firms from using the credit programs. Second, the available funds may have provided insurance value, even to firms that did not use the funds, in the same way that consumers get value from their insurance even in years when they do not make claims. Finkelstein argued that the second explanation of low take-up rates suggests free benefits from the programs rather than unfortunate frictions.

Deborah Lucas agreed that insurance value was an important consideration when discussing COVID-19-era policies, because unlike government programs, the insurance value of both fiscal policies and credit policies may have been more important than trying to stimulate demand.

Multiple participants raised questions about the political economy of the programs. John Sabelhaus asked whether there was any narrative evidence about what was occurring in legislative bodies while the programs were being enacted. Eric Zwick mentioned that credit programs for firms in particular often benefit policymakers or friends of policymakers, thus creating political risk.[2] Additionally, Zwick asked whether political economy could explain differences between countries' programs, highlighting differences in the financial sectors of the countries studied. Gee Hee Hong highlighted evidence that European countries relied on a common, agreed-upon framework while creating credit programs, which may help explain the similarities between programs across countries. Lucas additionally mentioned that in the United States budget scores that show the government making money are politically advantageous and are passed comparatively easily. She noted that the countries in the European Union had a less political approach to the credit programs than the Paycheck Protection Program in the United States.

Şebnem Kalemli-Özcan suggested bringing emerging market economies into the analysis. She mentioned that in her own work from 2021 with Pierre-Olivier Gourinchas, Veronika Penciakova, and Nick Sander,

2. Associated Press, "Congress Created PPP Loans—Then at Least 12 Members Reaped Their Benefits," *Fortune*, July 8, 2020, https://fortune.com/2020/07/08/ppp-loan-recipients-members-of-congress/.

small and medium-sized enterprise (SME) failures were reduced in both advanced economies and emerging markets.[3] Kalemli-Özcan argued that the prevention of SME failures had substantial macroeconomic effects. She added that while the fiscal and credit programs had small multiplier effects, they had large employment effects due to the prevention of SME failures. Sabelhaus noted similarly that while measures like GDP might capture downstream effects of these policies, other intermediate outputs, such as how many businesses kept their doors open, might help answer whether credit programs had their intended effects.

Lucas noted that the paper includes seven countries because there was insufficient data for other countries. She noted that while the small sample size limits the ability to draw inferences from the data, the relationship between fiscal programs, credit programs, and macroeconomic effects found among the seven countries points to the need for more work on the question. She also noted that the argument could be made that businesses' defaults were not prevented during COVID-19 but instead artificially lowered by the credit programs. She pointed to a recent uptick in defaults in Europe, which suggests that some problems were postponed rather than prevented as a result of the credit programs.[4]

Kristin Forbes asked whether there was any time series data regarding the composition of traditional fiscal versus off-budget loan guarantees and credit programs. She hypothesized that the proportion of off-budget programs may have increased over time and that in an era of high debt levels and fiscal constraints the proportion of off-budget programs may grow. Forbes also highlighted a figure in the paper showing that aggregate fiscal response was similar across countries but the composition of policies was different. She mentioned that her previous work with Katharina Bergant found that countries with higher debt levels were more likely to use a higher share of off-budget policies in their fiscal response, which aligns with the

3. Pierre-Olivier Gourinchas, Şebnem Kalemli-Özcan, Veronika Penciakova, and Nick Sander, "Fiscal Policy in the Age of COVID: Does It 'Get in All of the Cracks'?," working paper 29293 (Cambridge, Mass.: National Bureau of Economic Research, 2021), https://www.nber.org/papers/w29293; also in *Economic Policy Symposium Proceedings: Macroeconomic Policy in an Uneven Economy* (Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City, 2021).

4. Since this discussion took place, defaults in Europe have surged; Sudesh Baniya, "Business Bankruptcy in EU Surges to Highest Levels since 2015," Euronews, August 8, 2023, https://www.euronews.com/2023/08/26/business-bankruptcy-in-eu-surges-to-highest-levels-since-2015#:~:text=Business%20insolvencies%20in%20the%20EU,the%20food%20and%20housing%20sector.

findings of the paper by Lucas and Hong.[5] Lucas noted that she did not have access to time series data across countries while writing the paper.

Jason Furman asked how one would differentiate between the initial effects of credit programs and their downstream effects. Furman gave the example of bailouts that initially cost billions of dollars but also reduce the need for government programs like Medicaid, thus offsetting the cost. He also mentioned that government bailouts can increase GDP, which then increases tax revenue, again offsetting the cost of the initial bailout. He suggested drawing a conceptual line between the dynamic credit scoring in the paper and a complete dynamic model with second- and third-round effects.

Janice Eberly referred to a paper by Susan Cherry and coauthors from the Fall 2021 *BPEA* Conference as a partial explanation for why the incremental resources provided by forbearance appeared so low in the Hong and Lucas paper.[6] Eberly noted that most people who received forbearance on their mortgage payments returned to paying them off very quickly, and thus the size of the forbearance programs may look comparatively small against the fiscal policies and credit policies.[7]

Joe Beaulieu asked whether there were differences in efficiency between the types of policies that economists typically analyze, like taxes, and COVID-19-era fiscal and credit policies.

Lucas also responded to the paper's two discussants. In response to Alan Auerbach, Lucas discussed whether accrual accounting would be preferable in the budget. She said that in the United States budget there was a distinction drawn between firm legal commitments and commitments that could be changed ex post, with accrual only applying to a subset of the former. Lucas suggested that it is difficult to incorporate accrual accounting into a cash budget. She said that credit is currently accounted for on an accrual basis by capitalizing the projected net cash flows on loans made or guaranteed in a given year, in order to make the cost most comparable with

5. Katharina Bergant and Kristin Forbes, "Policy Packages and Policy Space: Lessons from COVID-19," *European Economic Review* 158 (2023):104499. https://www.sciencedirect.com/science/article/abs/pii/S0014292123001289.

6. Susan Cherry, Erica Jiang, Gregor Matvos, Tomasz Piskorski, and Amit Seru, "Government and Private Household Debt Relief during COVID-19," *Brookings Papers on Economic Activity* (Fall 2021): 141–99, https://www.brookings.edu/bpea-articles/government-and-private-household-debt-relief-during-covid-19/.

7. Juan M. Sánchez and Olivia Wilkinson, "Forbearance during COVID-19: How Many Borrowers Used It, and for How Long?," *On the Economy*, Federal Reserve Bank of St. Louis, May 31, 2022, https://www.stlouisfed.org/on-the-economy/2022/may/forbearance-covid19-how-many-borrowers-how-long.

items recorded on a cash basis. She commented further that while accrual accounting has clear advantages, it also risks budget discipline.

In response to Carmen Reinhart, Lucas acknowledged that if emerging markets were included in the analysis, an important further consideration would be international debt relief to those countries. Additionally, she agreed that forbearance was of special concern in emerging markets, where governments may not have paid for forbearance but rather instructed banks to grant forbearance. Finally, she agreed with Reinhart that these policies carry large risks, lack discipline, and therefore require attention. Hong agreed with Lucas and Reinhart that future work should focus on credit policies in emerging markets.

PINELOPI K. GOLDBERG
Yale University

TRISTAN REED
World Bank

Is the Global Economy Deglobalizing? If So, Why? And What Is Next?

ABSTRACT Data on global trade as well as capital and labor flows indicate a slowdown, but not reversal, of globalization since the 2008–2009 financial crisis. Yet profound changes in the policy environment and public sentiment in the largest economies over the past five years suggest the beginning of a new era. Increasing anxiety about the labor market effects of import competition from low-wage countries, especially China, laid the groundwork but was not the catalyst for the reversal in attitudes toward globalization. Similarly, the COVID-19 pandemic provided novel arguments against free trade based on global supply chain resilience, but neither the pandemic nor short-run policy response had enduring effects on trade flows. We demonstrate that global trade was remarkably resilient during the pandemic and that supply shortages would likely have been more severe in the absence of international trade. After a temporary decline in 2020, global trade in goods and services increased sharply in 2021. Russia's invasion of Ukraine raised new concerns about national security and the exposure of supply chains to geopolitical risk. This was followed by demands to diversify away from "non-friendly" countries and toward the employment of trade policy, export restrictions in particular, to halt China's technological development. The future of globalization is highly uncertain at this point, but these new policies will likely slow global growth, innovation, and poverty reduction even if they benefit certain industries in certain countries. Regarding resilience, the main goal of recent trade policy changes, measures of trade volatility or concentration can be helpful, but resilience will be elusive as long as we lack benchmarks against which policy performance can be measured.

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. The authors are not currently an officer, director, or board member of any organization with a financial or political interest in this paper.

Brookings Papers on Economic Activity, Spring 2023: 347–396 © 2024 The Brookings Institution.

At a December 2022 opening of a semiconductor chip plant in Phoenix, Arizona, Morris Chang, founder of Taiwan Semiconductor Manufacturing Company, stated "globalization is almost dead and free trade is almost dead . . . I don't think they will be back" (Ting-Fang 2022, par. 5). Such claims are not new. For the past decade economists have been debating the future of globalization, pointing out that since the financial crisis of 2008–2009, world trade has been growing more slowly than GDP—a reversal of an earlier trend observed during the two decades that marked the era of so-called hyper-globalization (1989–2009). Using data up to the onset of the COVID-19 pandemic, Goldberg (2019) and Antràs (2020) argued that there was little systematic evidence to support the view that the world economy had entered an era of deglobalization. However, the past three years have seen dramatic changes in trade policy and geopolitical environments that call for a reevaluation of this view.

No matter how one feels about globalization these days, there is wide consensus that it has had substantial effects on global growth, poverty reduction, and inequality (both across and within countries)—not to mention its political, societal, and cultural consequences.[1] It is no surprise that the possibility of its reversal has become a central question in policy and public discourse. This paper has three aims: (1) to critically assess existing evidence regarding the deglobalization hypothesis, (2) to analyze the causes of a potential deglobalization trend, and (3) to assess, speculatively at this point, the consequences of potential deglobalization. It is accordingly organized in three sections, each reflecting these aims.

Traditional metrics of globalization (trade, capital flows, immigration) still show no sign of trend reversal—if anything, they suggest that trade has rebounded after the COVID-19 pandemic. However, the policy environment and public sentiment toward globalization have fundamentally changed, especially in the largest economies. To the extent that policy and public opinion help shape economic outcomes, there are therefore good reasons to believe that we have entered a new era. Regarding the causes of this trend, there seem to be three phases in the deglobalization sentiment.

The first phase starts around 2015 with concerns about the impact of import competition from low-wage countries, especially China, on the labor market, and the impact of refugee flows, primarily in Europe. Such concerns

1. See Irwin (2019) and Chari, Henry, and Reyes (2021) for a review of the evidence on the relationship between trade and growth and a reevaluation of the critical assessment of Rodríguez and Rodrik (2000).

make the general public receptive to protectionist policies. This phase is marked by Brexit and the United States and China increasing tariffs on one another, the economic effects of which were meaningful but still not substantial enough to reverse decades-old globalization trends.

The second phase plays out during the COVID-19 pandemic when new arguments against trade emerge: temporary shortages of personal protective equipment and other items are attributed to the fragility of global supply chains. Demands for greater resilience through greater dependence on domestic production provide a novel justification for reshoring economic activity. We argue that the evidence does not support these claims and that, if anything, trade increased economies' resilience during the COVID-19 pandemic. While the demands for resilience did not have any direct implications for policy or trade during this period, and trade actually grew fast following the pandemic, the extensive coverage of the topic in the press further contributed to the notion that globalization had harmful effects to the domestic economy and prepared the ground for the third phase.

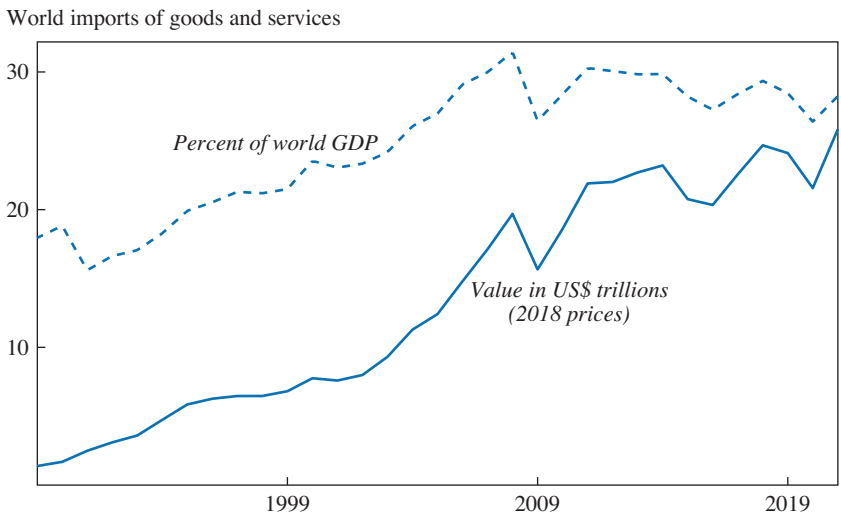
The third phase begins with the Russian invasion of Ukraine in February 2022 and provides the perhaps strongest argument to date for rethinking globalization: concerns about national security. The reliance of Europe on Russia for energy exposes the fragility of a global supply system based on hyper-specialization. New demands for "decoupling" emerge, not just from Russia, but from any country that is not our friend, and a new term enters the international trade vocabulary: "friendshoring." The talk is followed by strong policy actions by the United States that include sweeping export restrictions in the semiconductor sector targeting China (McKinnon and Fitch 2022; Ellerbeck 2023). These developments can plausibly be considered the markers of a new era.

Given that these trends are only a few months old, any assessment of their consequences is inevitably highly speculative. With this caveat in mind, we offer some concluding thoughts and very preliminary evidence on possible effects of deglobalization on efficiency, inflation, global and within-country inequality, resilience, and peace.

I. Is the World Economy Deglobalizing?

While we do not formally define globalization, we employ three different approaches to capturing trends toward or away from it. First, we examine trends in some key variables (such as world exports and imports, capital flows, and immigration) that have been traditionally viewed as appropriate metrics of globalization. Second, we assess the trade policy environment.

Figure 1. Trade Is Growing, but Has Declined as a Percentage of World GDP since the Global Financial Crisis



Sources: COMTRADE; World Trade Organization; and World Bank.

Note: Nominal value of goods imports is from COMTRADE, and nominal value of services trade is from the World Trade Organization, both reported in US dollars. The sum of nominal trade values is divided by world GDP in US dollars at market exchange rates (NY.GDP.MKTP.CD GDP in the World Bank Development Indicators). Nominal trade values are converted to 2018 prices using the GDP deflator (NY.GDP.DEFL.ZS) for the United States.

Finally, we consider how public discourse has changed in recent years by analyzing the occurrence of certain terms in news articles.

I.A. Trends in Standard Metrics of Globalization

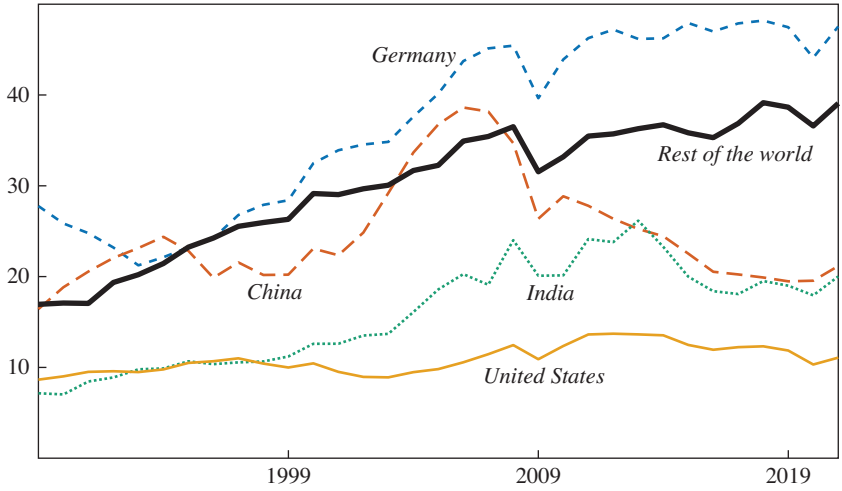
A popular measure of globalization is international trade in goods and services. World imports have grown rapidly in the last three decades, including post-2009, as shown in figure 1. Notably, world imports increased sharply in 2021 after a temporary decline in 2020 due to COVID-19, showing trade's longer-term resilience to the pandemic. Yet, when measured as share of GDP, imports have slightly declined since the global financial crisis. This is the aforementioned trend that has led to concerns that the world has started deglobalizing post-2009. Because the decline is too small to justify the term "deglobalization," "slowbalization" has been used instead as a more appropriate characterization of this trend.

The mirror of world imports are world exports, which generate an almost identical picture. Panel A of figure 2 reports trends in exports since 1989

Figure 2. Countries Have Diverse Experiences with Globalization

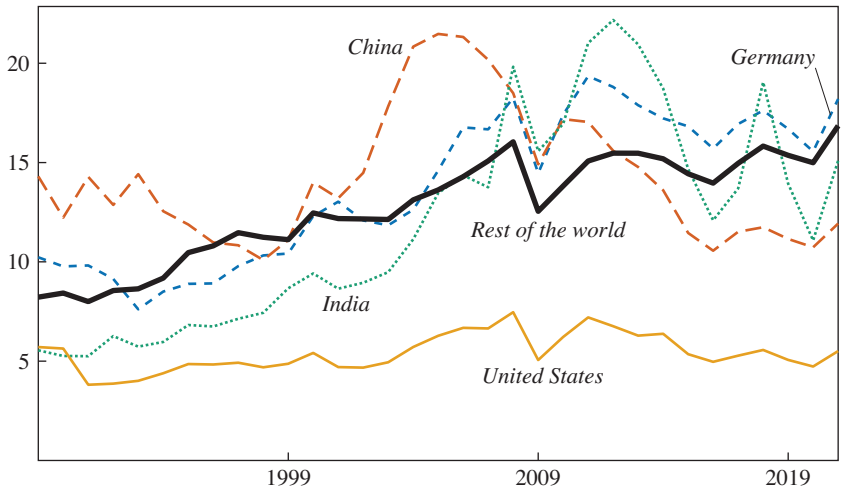
Panel A: Exports of goods and services

Percent of country's GDP



Panel B: Intermediate goods imports

Percent of country's GDP

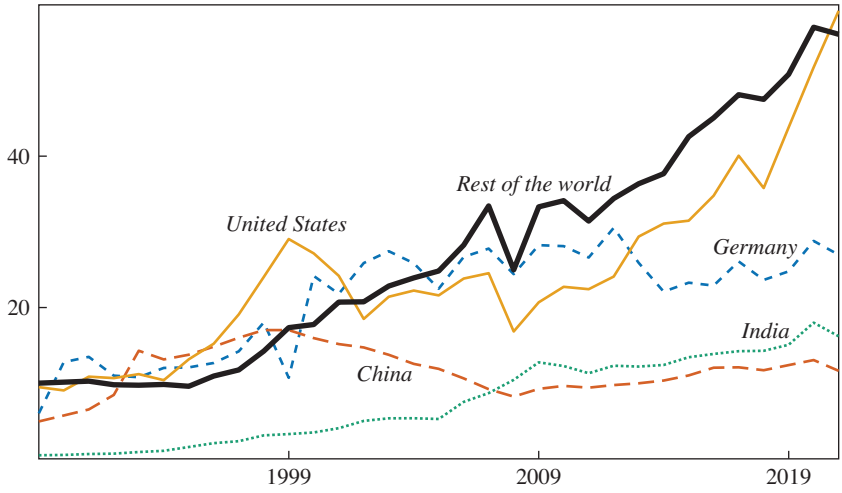


(*continued on next page*)

Figure 2. Countries Have Diverse Experiences with Globalization (*Continued*)

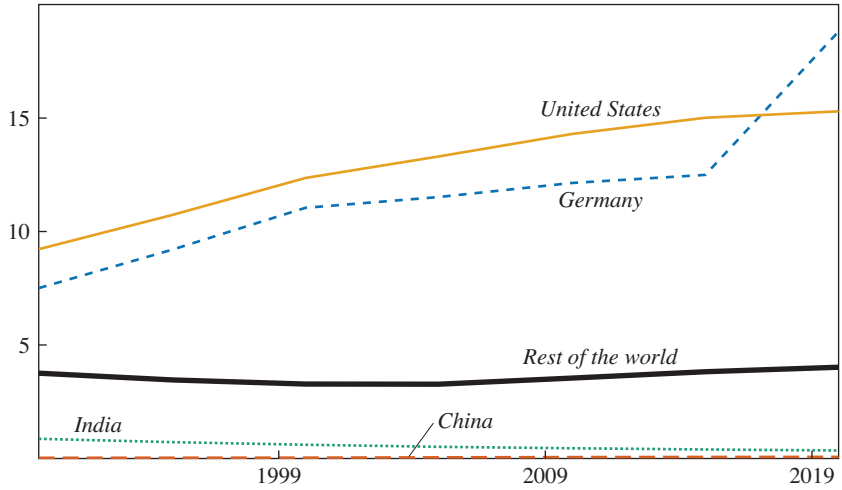
Panel C: Stock of inward foreign direct investment

Percent of country's GDP



Panel D: Stock of inward migrants

Percent of country's population



Sources: COMTRADE; United Nations Conference on Trade and Development (UNCTAD); United Nations Global Migration Database; and World Bank.

Note: Intermediate goods in panel B are classified according to the Broad Economic Categories (BEC4) of the Standard International Trade Classification system. The stock of foreign direct investment in panel C includes equity investments with foreign ownership greater than 10 percent and reinvested profits accruing to that ownership. The stock of inward migrants in panel D includes all foreign-born residents of a country.

for four large economies and the rest of the world. Examining these trends at the individual country level highlights how amid global trends, national integration with the global economy is changing. China and India appear to be less reliant on the global economy. Exports as a share of GDP have declined in both countries, from peaks during the 2000s in the case of China and the 2010s in the case of India, leveling off at about 20 percent of GDP in both economies. On the other hand, Germany's exports have been increasing since the 1990s, except for brief interruptions during the global financial crisis and COVID-19 pandemic, so that exports now account for almost 50 percent of GDP. In the United States, exports have held steady at about 10 percent of GDP. In the rest of the world, exports are about 35 percent of GDP, indicating greater reliance on the global economy for demand than China, India, or the United States.

These figures highlight diverse experiences with globalization. Globalization has allowed the United States to enjoy the consumption of foreign goods, borrowing to purchase the exports of the rest of the world. China and India have exploited this demand to fuel their own growth from a low income. They now rely less on the global economy and instead on their own populations, which are richer and can support domestic demand. Germany, which is nearly as rich as the United States, shows low wages are not a prerequisite for export success.

The imports of intermediate goods and services as a share of countries' GDP are shown in panel B of figure 2. Such imports are essential inputs into exports, especially for countries participating in global value chains (GVCs), where different stages of production processes are located across different countries. The World Bank (2020) estimates that GVCs account for 50 percent of global trade but that their growth has plateaued. Intermediate goods imports are a declining share of GDP in China and India, as both countries now produce more inputs domestically. These two economies are becoming less reliant on the global economy for inputs. Though it is less stark, there is also a downward trend in intermediate imports in Germany and the United States. In the rest of the world, the trend is instead slightly upward. Overall, the trends in the trade data suggest a slowdown of global trade, but this is hardly surprising given the extraordinarily fast trade growth during the two decades of hyper-globalization (1989–2009).[2]

Measures of globalization in factor markets provide a different perspective. The stock of inward foreign direct investment (FDI) in panel C of figure 2 measures the globalization of capital markets. In the United States

2. For a more detailed discussion of this point, see Goldberg (2019) and Antràs (2020).

and the rest of the world the stock of inward FDI accounts is valued at nearly 60 percent of GDP and shows no major downward trend. In Germany, inward FDI as a share of GDP in the last two decades grew until the global financial crisis but has not yet recovered from that peak. China experienced a boom in FDI in the 1990s that preceded the boom in exports in the 2000s. In India, the rise in exports and increase in FDI from the mid-2000s onward go hand in hand. Figure A1 in the online appendix shows positive trends are similar when counting the sum of debt and equity foreign investment, using the series of Coppola and others (2021), which also incorporates foreign investment through tax havens. In panel C of figure 2, the level of inward foreign investment is greater in China and Germany relative to the level in figure A1 in the online appendix, reflecting greater reliance on debt financing in these economies. The series in figure A1 includes debt in addition to foreign direct investment, which includes only purchases of equity shares.

The stock of inward migrants in panel D of figure 2 measures the globalization of labor markets. Germany and the United States have absorbed migrants as an increasing share of their population, with Germany recently surpassing the United States as it absorbed a surge of refugees in the 2010s. China and India are home to far fewer migrants as a share of their population, and this trend appears flat.

There are two main takeaways from figure 2. First, it is clear that the growth of trade as a percentage of GDP has stalled since the financial crisis, and even declined in some cases. However, trends in capital and labor markets tell a different story. Taken together, these trends suggest that it is premature to talk of deglobalization—the slowdown of global trade seems a natural development following its earlier fast growth and reflects partly the growth of the domestic markets of two large, formerly low-income countries, China and India. Second, deglobalization trends are highly heterogeneous across countries. While the United States and China—the world's two largest economies, which by virtue of their economic size drive aggregate trends—seem to be gradually decreasing their reliance on global markets, this is not true for the rest of the world.

I.B. Policy Environment

While the data do not yet show strong signs of deglobalization, the trade policy environment has changed drastically in the past five years, especially in the United States. This is important as trade and other measures of globalization may respond with a lag to changing policies.

Since the end of World War II, global trade barriers had been gradually falling, while a proliferation of bilateral and regional trade agreements allowed many countries to become more integrated in world markets. Several developing countries reduced their tariffs unilaterally and joined the World Trade Organization (WTO), which allowed them to enjoy many of the benefits of multilateralism. The United States and other advanced economies played a leading role in the reduction of trade barriers and the design of the world trading system.

The picture changed dramatically in 2018, when the United States announced a first set of tariff increases targeting several countries, but especially China. Eventually these tariff increases led to a tariff war between the United States and China, the world's two largest economies in 2018 and 2019. The United States also imposed tariff increases on steel and aluminum imports from nearly all countries. Despite the change in the US administration, most of these tariffs remain in place to date. The United States continues to argue within the dispute resolution process that these tariffs are required for national security, though it has also declined to elaborate the mechanism through which discriminatory tariffs improve its national security. In December 2022, in response to challenges brought by China, Norway, Switzerland, and Türkiye, a WTO dispute settlement panel found that the tariffs indeed violated WTO rules. Article 21 of the General Agreement on Tariffs and Trade allows WTO members to discriminate against one another for national security reasons "in time of war or other emergency in international relations" (21.b.3) but the panel concluded that the year 2018 did not constitute such a time.

Additionally, the US practice of exercising its veto power to block the appointment of new judges in the WTO's Appellate Body as the terms of existing judges ended has rendered the WTO's dispute settlement system defunct. US dissatisfaction with the way trade disputes between countries are resolved within the international rules-based system had been simmering for years, but it was not until the end of 2019, when the terms of the last two judges expired, that the crisis reached its peak, paralyzing the dispute settlement process, effectively ending twenty-three years of WTO trial-like enforcement, the keystone of international efforts to prevent trade protectionism. The Appellate Body now lacks a quorum to hear appeals, so that the United States, for instance, could not even appeal this recent ruling if it wanted to. One interpretation of US actions is that they are less about dissatisfaction with the WTO system for resolving disputes than about dissatisfaction with outcomes that are not in the United States' favor.

While other countries could take this approach, some nonetheless still find the spirit of WTO dispute resolution useful, with, for instance, the EU and Türkiye recently opting to use arbitration in the absence of the Appellate Body to resolve a dispute over rules requiring local content in pharmaceuticals purchased by Türkiye's national health scheme.[3] The EU challenged these rules and won in arbitration, an alternative mechanism available in the WTO charter to appeal disputes. Although Türkiye lost, with the arbitrator finding WTO rules require it to end a "buy Turkish" rule for pharmaceuticals, Türkiye was willing to go through arbitration.

More recently, there has been a clear shift in the US approach to trade away from liberalization and multilateralism toward industrial policy and discriminatory policies vis-à-vis China in the name of supply chain resilience and national security. This shift was first apparent in the speech that Ambassador Katherine Tai, the US Trade Representative, gave at the Roosevelt Institute's Progressive Industrial Policy Conference on October 7, 2022.[4] It became even more salient a week later, when the White House released its *National Security Strategy*, which poignantly begins with the words: "The world is changing" (White House 2022, 6). The report came a few days after the announcement of sweeping export restrictions on the semiconductor industry aimed at stopping China from advancing technologically. According to the report, the restrictions were justified in the name of national security, specifically, the military-civilian fusion as practiced by China, and the increasing significance of dual-use goods, that is, goods that are designed for commercial applications but that have military applications.

These actions signal a clear break from former US trade policy and the possible beginning of a new era. So does the CHIPS and Science Act which provides multibillion-dollar support for the development of the domestic semiconductor industry, several discriminatory provisions in the Inflation Reduction Act toward countries that are not members of the United States-Mexico-Canada Agreement, and the United States distancing from new trade agreements, such as the Comprehensive and Progressive Agreement

3. World Trade Organization, "Turkey: Certain Measures concerning the Production, Importation, and Marketing of Pharmaceutical Products," https://www.wto.org/english/tratop_e/dispu_e/cases_e/ds583_e.htm.

4. Office of the United States Trade Representative, "Remarks by Ambassador Katherine Tai at the Roosevelt Institute's Progressive Industrial Policy Conference," https://ustr.gov/about-us/policy-offices/press-office/speeches-and-remarks/2022/october/remarks-ambassador-katherine-tai-roosevelt-institutes-progressive-industrial-policy-conference#:~:text=We%20believe%20industrial%20policy%20and,opening%2C%20liberalization%2C%20and%20efficiency.

for Trans-Pacific Partnership (CPTPP). The United States did in May 2022 propose a new economic agreement with the largest economies in Asia and Oceania except for China, called Indo-Pacific Economic Framework for Prosperity (IPEF). This agreement is envisioned to include commitments to standards for digital trade, decarbonization, labor standards, and tax enforcement, but there has been no mention of US market access, which has been the traditional focus of trade agreements, and which is also highly valued by trade partners.[5]

Meanwhile, several other countries are going in the opposite direction, at least when it comes to economic agreements. The recent signing of the Regional Comprehensive Economic Partnership between the ten member states of the Association of Southeast Asian Nations that went into effect on January 1, 2022, the expanding membership of the CPTPP as more countries request to join, and the launching of the African Continental Free Trade Area which aims to boost intra-African trade, are examples of a trend toward more regional or—in the case of the CPTPP—plurilateral integration. Along the same lines, Fajgelbaum and others (2022) show that while the trade war between the United States and China reduced trade in the targeted product categories between these two countries, it boosted trade between other countries and the rest of the world in these products. The countries that exhibited the largest export growth to the rest of the world were countries with expansive trade agreements. So the trade conflict between the world's two largest economies did not simply cause reallocation of global trade flows, it also generated new trade opportunities for other countries. Therefore, it seems that outside the United States the picture is more nuanced, with many countries striving to take advantage of the new opportunities potentially created by the reversal of US policy.

Note, however, that one of the most assertive recent moves of the United States, the export restrictions in the semiconductor sector targeting China, requires the cooperation of other countries to succeed. Because most of the semiconductor chip manufacturing takes place in countries other than the United States—using, however, to a large extent US software and US machinery—the new export controls require third-country chip manufacturers to obtain export licenses from the United States in order to export their products to China. The United States implemented this policy change

5. The CPTPP did include commitments on digital trade, environment, and labor which the United States had lobbied for before it exited the agreement. From this perspective the IPEF could be a way for the United States to achieve some of the goals it had pursued in the CPTPP negotiations, but without making commitments of further market access.

unilaterally, putting allies in a difficult position as they faced a choice between going along with the export restrictions or—if they wished to continue exporting to China—forgoing access to US technology. This use by the United States of economic interdependence to force other countries to go along with its policies has been called "weaponizing interdependence."[6] It suggests that no matter what other countries' preferences and intentions are, ultimately it is the priorities of the largest economy of the world that may shape the future of the global economic system. Seen in this light, the United States turning inward may have important implications for the future of globalization.

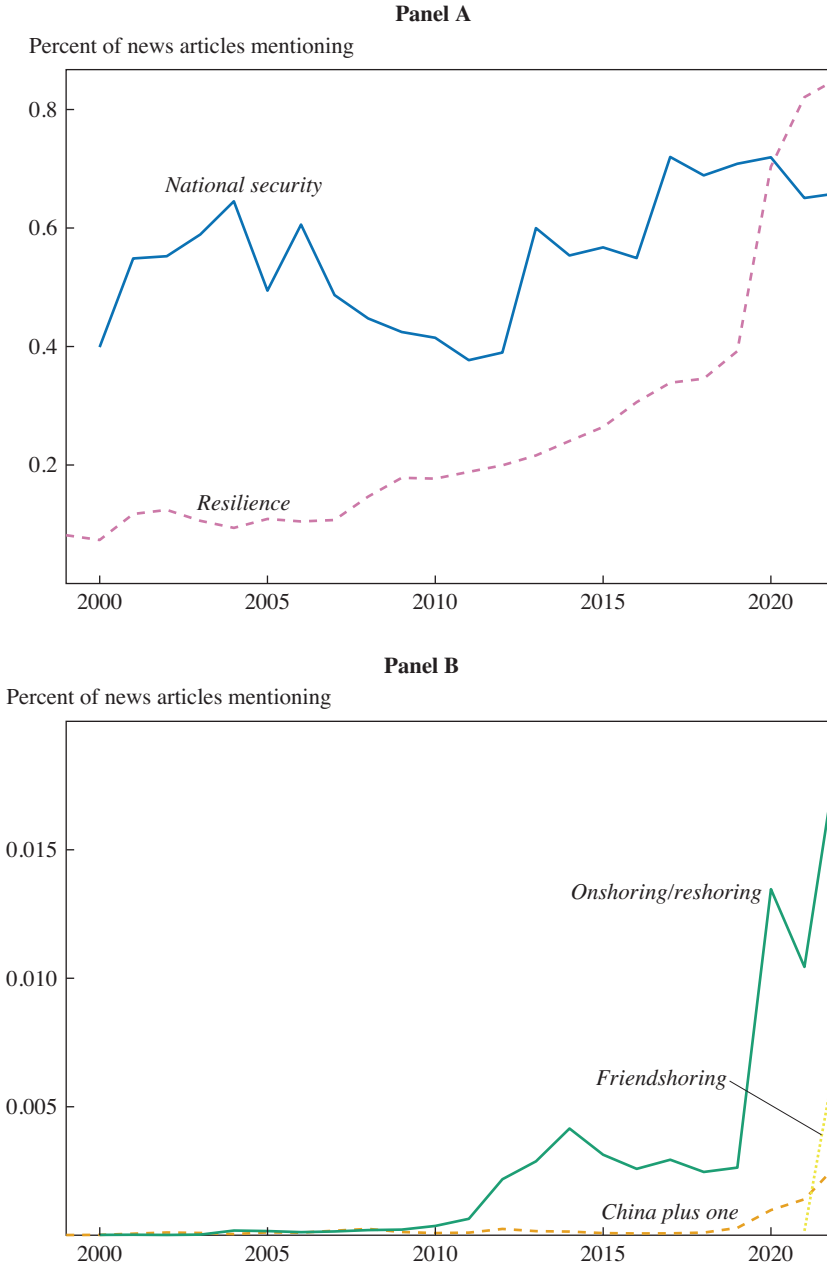
I.C. Public Discourse

Just as economic variables respond with a lag to policy, policy responds with a lag to public sentiment and attitudes. Therefore, we now turn to an investigation of such attitudes regarding trade and globalization. Data from the PEW global attitudes surveys show that the public still views trade as beneficial to the economy.[7] However, news articles suggest more nuanced attitudes marked by increasing skepticism about participation in the global economy.

Panel A of figure 3 shows that mentions of the term "resilience" have been increasing since the global financial crisis, with an acceleration during the COVID-19 pandemic. This could reflect an increasing awareness that global supply chains expose countries and companies to risk but need not cast doubt on the underlying trade. Doubts may instead be reflected in the increasing use of the phrase "national security." Surprisingly, today this phrase occurs in a greater percentage of news articles than immediately after September 11, 2001, and after the Arab Spring and the NATO intervention in Libya in 2011. It is difficult to pinpoint the proximate cause of the increasing salience of national security in 2013 and again in 2017. The beginning of the nationalist Donald Trump administration is one hypothesis for the latter, and though mentions of national security dropped sharply after the election of Joe Biden in 2021, it is still used nearly twice as much as in 2000. Panel B of figure 3 shows the emergence of other less common jargon. The words "onshoring" and "reshoring" emerged into frequent use in 2010. This suggests that skepticism of globalization predated the Trump administration, during which use of the terms skyrocketed.

6. The phrase was coined by Farrell and Newman (2019).
7. See Dorn and Levell (2021, 32, fig. 8).

Figure 3. Trends in Documents Mentioning "Resilience," "National Security," "Onshoring" or "Reshoring," "Friendshoring," "China Plus One"



Source: Factiva.

Not all discourse has been nationalist. After 2017, the phrase "China Plus One" also began to emerge in the business press, a recognition of China's important role in global supply chains and the need to diversify to at least one alternative supplier. In 2021, US Secretary of Commerce Gina Raimondo introduced the concept of friendshoring.[8] The term is now associated with policy aimed at moving US value chains away from China. Yet, compared to onshoring or reshoring, it also suggests a more moderate view of globalization, that it should continue and may deepen, but only with a subset of countries who are friends.[9]

I.D. Summary

Trends in goods and services trade, both for the global economy and for major economies, suggest a slowdown (though not reversal) of globalization. But FDI and migrant flows suggest the opposite. Overall, to date, there is no hard evidence in the data that we have entered a new era. There are, however, profound changes in the policy environment, at least in the United States, as well as in general attitudes toward trade with other countries that may be the harbinger of—if not deglobalization—a new kind of globalization. Prominent among these are the concern about national security, the demand to diversify away from China, and the belief that trade should only be promoted if it is between friends. These issues are distinct from the more traditional concern about the impact of globalization on low-skill workers in advanced economies, which has traditionally been the reason for advocating reshoring. Their origins seem rooted more in (geo) politics than economics. Accordingly, we may be entering an era where the future of trade and globalization is shaped top-down by politically motivated governments rather than by market forces.

II. Causes

II.A. What Caused Hyper-Globalization?

The factors that have contributed to the hyper-globalization of the 1990s and 2000s have been extensively analyzed in the literature. The World

8. US Department of State, "Telephonic Press Briefing with U.S. Secretary of Commerce Gina M. Raimondo," November 18, 2021, https://www.state.gov/telephonic-press-briefing-with-u-s-secretary-of-commerce-gina-m-raimondo/.

9. This stance is also reflected in US Secretary of the Treasury Janet Yellen's (2022) piece on resilient trade.

Bank's 2020 *World Development Report* on global value chains (GVCs) provides a succinct summary of the main drivers, which involve a combination of technology and policy. Major technological advances after World War II led to dramatic declines in transportation and communication costs that enabled the fragmentation of the production process, so that different stages of production could be outsourced to different countries to take advantage of international cost differences. This process led to the emergence and growth of GVCs and to unprecedented growth in gross trade as intermediate goods and services crossed borders many times before final products reached their intended destinations. But technology alone would not have achieved this growth had it not been for a set of policies that created the right environment for trade and GVCs to flourish: sharp reductions in tariff and nontariff barriers, often within the context of unilateral trade liberalizations by developing countries; bilateral and regional trade agreements that liberalized trade for the member countries; and last but not least, the expansion of the WTO to include several developing countries, including China.[10] Multilateralism played an important role during this period by reducing uncertainty. When goods and services are traded across the entire globe, regional agreements do not suffice; only multilateralism can offer the degree of stability and predictability required for global trade to flourish.[11] At the same time, the ideological shift toward neoliberalism, fueled by the fall of communism in Eastern Europe and the former Soviet Union, provided a supporting intellectual backdrop for the changes that were taking place, encouraging the entry of several countries in Eastern Europe, Central Asia, and East Asia into the global trading system.

Against this background, it is natural to ask what has changed since that time, causing a possible reversal of earlier trends. Not surprisingly, the main drivers are the same factors that led to the hyper-globalization of the past: technology and policy.

II.B. What Is Causing a Possible Reversal?

TECHNOLOGY There are two reasons one might suspect that the slowdown of trade (as percentage of GDP) since the financial crisis could be technologically driven.

10. See Irwin (2022) for an overview of the trade reform wave of 1985–1995, the years leading up to the era of hyper-globalization.

11. For a more detailed discussion, see Goldberg and Pavcnik (2016), Goldberg (2019), and Goldberg and Larson (2023).

The first reason is that the fragmentation of production underlying the expansion of GVCs and trade would eventually reach its technological limit. In fact, some have viewed the decline of trade in intermediates (evident in figure 2) as evidence that the fragmentation of production has run its course. However, as Goldberg (2019) argues, this evidence is far from conclusive. Apart from the fact that the decline in intermediates trade is too small to support such a conjecture, the *value* of trade in intermediates is influenced by several factors, including commodity prices. An alternative measure of fragmentation that is more closely associated with GVC trade, the share of parts and components in *volume* terms in manufacturing trade, has increased at a moderate pace since the 1990s and has not shown any signs of reversal since the global crisis (Goldberg 2019). Accordingly, the evidence on this point is mixed.

The second argument is that recent technological advances, for example, automation and 3D fabrication, now favor reshoring of economic activity as the production process relies less on low-skill labor than in the past. However, this argument does not withstand scrutiny for three reasons.

First, as the World Bank's (2020) report showed, recent technological advances also create additional scope for trade. Indeed, the evidence to date suggests that those sectors most affected by automation (e.g., the automobile industry) expanded intermediate product imports, especially from developing countries, as the scale of production increased.

Second, as Antràs (2020) has argued, the presence of sunk costs in foreign investments implies that reshoring is not symmetrical to offshoring; even if the incentives to produce abroad may not be as strong today as they were in the 1990s, GVCs will not abandon their activities in other countries given the large sunk costs they have incurred in the past.

Finally, it is conceivable that the advances in modern technology, especially in information and communication technology, lead us in the opposite direction, fostering another wave of globalization, this time driven by trade in services.[12] The increase in remote work and internet-based services and commerce during COVID-19 has contributed to this trend. However, trade in services requires further global integration and regulatory

12. This view is advanced in Baldwin (2019), for instance. Trade in services has indeed grown quickly in recent years, and there is certainly scope for further growth in the future. However, it is important to distinguish between level and direction in this context—in terms of level, there is limited trade in services (compared to goods) at present, even within highly integrated areas such as the European Union. We note also that measuring trade in services poses formidable challenges, so that trends in the data should be interpreted with caution.

convergence across countries, for which there is no appetite now.[13] Which brings us to policy.

POLICY Given that there is no compelling evidence to date that the trend against globalization is technologically driven, policy emerges as the primary explanation. In fact, that set of policies, discussed in section I.B, suggest that the policy environment has changed dramatically since the 1990s and 2000s. But this leads to another question: If deglobalization trends are driven by policy, what explains the change in policy?

Policy in all countries, but especially in democracies, responds to public sentiment. As discussed in the first section, the sentiment vis-à-vis globalization has been gradually changing in the United States since the mid-2010s. There are several reasons responsible for this change that are explored in the next section.

II.C. Phases of Backlash and Causes

There are roughly three phases in the backlash against globalization. The first phase starts in the mid-2010s, when politicians both at the extreme right and extreme left blamed international trade, in particular NAFTA and increasing import penetration by China, for the decline in manufacturing employment and stagnation of real wages in the United States. The sentiment that globalization had become a liability for the domestic labor market was not confined to the United States, as evidenced by Brexit in the United Kingdom and the strong anti-immigration movement in several other European countries. The second phase coincides with the unfolding of the COVID-19 pandemic (2020–2022), which led to additional concerns about trade related to the resilience of global supply chains. The third phase began with the Russian invasion of Ukraine in February 2022 and is centered on yet another concern: national security. The factors contributing to the anti-globalization sentiment in each phase are accordingly specific to each phase.

FINANCIAL CRISIS Given that the financial crisis marks the onset of the trade growth slowdown, it is tempting to attribute the sentiment against

13. For a detailed discussion, see, for instance, Mattoo (2018). The provision of many services requires licensing, but licensing requirements are not harmonized across countries. Even business services, such as banking and insurance, that can be purchased from firms in other countries, often raise delicate regulatory issues, especially regarding the handling of data and privacy. The challenge for policy is to strike a balance between the legitimate use of domestic regulation to protect consumers and its protectionist abuse. Trade policy alone is not enough to make progress in these areas; one also needs regulatory cooperation and coordination.

global markets to the crisis. Indeed, people in different countries were exposed to risks not of their own making that were "imported" from the United States. However, there are several reasons arguing against the hypothesis that the financial crisis was the origin of the anti-globalization sentiment. First, there is the timing. International trade declined sharply during 2008–2009 but recovered immediately afterward. It is not until 2015–2016 that strong arguments against trade and globalization emerge. Of course, one could argue that it took time for people's frustration to build up, but five to six years seems an implausibly long lag. Second, several countries, most importantly China and some in the developing world, fared well during the crisis in terms of growth and poverty reduction, so that the pain felt was by no means universal. For the United States and Europe, Bonomi, Gennaioli, and Tabellini (2021) provide an interesting explanation of why the financial crisis did not lead to a fundamental change in attitudes (including those toward globalization).[14] They argue that it is not economic shocks per se that lead to changes in social attitudes but the interaction of economic shocks with a preexisting divide along cultural lines. The financial crisis may have caused economic hardship, but this hardship was not concentrated in low-education, socially conservative groups that had a preexisting aversion to globalism. However, this was not the case with the next possible cause of the backlash.

IMPORT COMPETITION FROM LOW-WAGE COUNTRIES, ESPECIALLY CHINA As noted earlier, starting in the mid-2010s, several US politicians blamed increasing inequality in the United States, declining manufacturing employment, and stagnation of real wages on imports from low-wage countries, especially from China—the so-called China Shock or China Syndrome. While there is little evidence to support the view that such imports had a major effect on aggregate trends and outcomes in the United States, there is ample evidence by now that they had a large impact on local communities that were disproportionately affected by import competition.[15] These communities fit the profile in Bonomi, Gennaioli, and Tabellini (2021): they consisted of low-income, low-education, socially conservative households who, faced with a major economic shock, became polarized and increasingly economically conservative, rejecting redistribution as well as globalization.

14. This point is also made by Guido Tabellini in his Presidential Address to the Econometric Society in 2022; https://www.youtube.com/watch?v=i6Cp1MIv-Rg&t=2s.

15. See Dorn and Levell (2021) for a discussion of the evidence for the United States and several European countries.

While the change in attitudes was initially confined to these communities, it garnered momentum over time, with trade and globalization becoming major issues in the 2016 electoral campaign. Eventually, this led to the US-China trade war of 2018–2019, a major departure from the earlier trade liberalization policies practiced by the United States. However, even as the trade war was unfolding, it seemed unlikely that this was a permanent change of policy; rather, the policy toward trade and China seemed specific to the choices of a particular administration at the time. Many commentators expected the trade restrictive policies to be reversed once a new administration came to power. This expectation did not materialize, indicating that the change in attitudes toward globalization was more permanent.

CLIMATE CHANGE AND CARBON TAXES With the emergence of climate change as a major challenge facing the world today, globalization has come under fire as a potential contributor to high emissions. The multiple crisscrossing of goods across borders as they are traded within GVCs implies additional packaging and fuel for transportation. Countries have different environmental standards, which may create incentives for pollution havens (though to date there is no evidence supporting this hypothesis).[16] Last but not least, trade is associated with growth, and growth (at least to date) means more pollution. Policies to cope with climate change, including carbon border adjustment taxes, have the potential to lead to a new world order as they will change relative prices with potential implications for countries' competitiveness and comparative advantage. However, the war in Ukraine and the associated energy crisis have put such adjustments on hold. Hence, while climate change could be a major force for rethinking transnational relationships in the future, it has not played this role to date.

COVID-19 AND DEMANDS FOR RESILIENCE Since the onset of the pandemic, a rather novel concern regarding trade has emerged: the resilience of global value or supply chains. Short-run supply shortages of various items, from paper towels and toilet paper to personal protective equipment and ventilators in March 2020 were attributed to the disruption of the normal functioning of GVCs due to COVID-19. Such concerns became even more pronounced toward the end of 2020 and in early 2021, when problems with shipping, delays at ports, and shortages of critical products such as baby formula were featured daily in the press. A chain with multiple links, some of which may be located in different countries, is as strong as the weakest link, some

16. See Shapiro and Walker (2018).

argued. Any time a link in a foreign country breaks due to a local shock, the global supply chain suffers. As a result, the idea that a natural way to increase the resilience of supply chains was to relocate as many links as possible to the domestic economy, that is, to reshore supply chains, gained traction during that time.

This was not the first time that vulnerability of GVCs to local shocks became a cause for concern. Boehm, Flaaen, and Pandalai-Nayar (2019) and Carvalho and others (2021) studied the 2011 earthquake in Japan and showed how this country-specific shock propagated through supply networks to affect firms located in other countries. Barrot and Sauvagnat (2016) studied firm-level idiosyncratic shocks due to natural disasters and found that such shocks affected not only the companies nearby within the region but also their customers who reported declines in sales growth in the medium run. However, neither the Japanese earthquake nor natural disasters led to the sustained criticism of GVCs and denouncing of globalization observed during the pandemic. This is surprising, especially in light of the fact that the COVID-19 shock was not country-specific, but global, weakening the case for reshoring. A natural question then is whether the resilience argument reflects genuine worry as opposed to presenting a novel and convenient excuse for pursuing old-style protectionism motivated by the aforementioned concerns regarding the labor market effects of trade and their potential contribution to domestic inequality.

To address this question, it is necessary to have a well-defined notion of resilience. A useful starting point is the definition provided by Markus Brunnermeier in his recent book, *The Resilient Society*. According to Brunnermeier, resilient means to "bend but not break" in response to a shock (2021, 2). A comparison between the oak and the reed is illuminating. The oak is robust and can withstand many shocks. But if the shock is strong enough, the oak may break. In contrast, the reed is not robust, but it is resilient. Even a light breeze makes it sway back and forth, but when hit with a big storm, it bends, it does not break, and proves therefore more resilient than the oak.

Illuminating as this comparison may be, it still begs the question of how one would operationalize the concept of resilience in economics. A strict interpretation of the "bend but not break" criterion would imply that an economic entity—be it a firm, an industry, an economic relationship, or an entire economy—would be resilient if it survived an economic shock and not resilient if it ceased to exist. A less strict interpretation might be taken to imply that even entities that survive a shock are not resilient if it takes

them a long time to completely recover. But this in turn begs the question of how long is a long time? It is becoming apparent that the concept of resilience raises several questions that remain unanswered in economics to date. Importantly, we lack a well-defined benchmark of resilience by which responses to economic shocks can be judged.

This state of affairs notwithstanding, one can make progress by making explicit some key factors that need to be considered when judging responses to economic shocks. The following outlines relevant considerations:

- Nature and magnitude of shock:
  - Supply, demand, or both[17]
  - Sector-specific, country-specific, or both
  - Idiosyncratic or systemic
- Time horizon (short, medium, or long-run)
  - Dependent on sector (e.g., food, medicines, where time is of the essence)
  - Dependent on (possibly non-homothetic) preferences (e.g., consumers in rich countries without well-developed public transportation may consider a car a necessity)
- Level of aggregation
  - Economy
  - Industry
  - Firm
  - Household

To see why these factors matter, consider first the response to a demand-side shock, for example, a decline in aggregate demand due to rising unemployment or due to health concerns. Whether or not the supply chain is national or global, or well diversified, will matter little for resilience in this case. Now consider the response to a supply shock, for example, a natural disaster or a disease. If the shock is specific to a particular location or country, then diversification of the supply chain to include multiple locations will make it more resilient. On the other hand, if supply shocks are correlated across locations (as is the case with a pandemic, for example), then diversification will do little to increase the supply chain's resilience. Along the same lines, if a supply shock originates in the domestic economy, then reshoring of supply chains will make them less resilient to such a shock. Of course, the opposite would be true if the shock originated abroad. The magnitude of the shock and

17. In practice, it may be hard to separate supply and demand shocks. Guerrieri and others (2020) show how a shock that starts as a supply shock can become a demand shock (COVID-19 is a good example).

relevant time horizon also matter greatly for resilience. For example, the 2011 earthquake in Japan caused severe disruptions in the global automobile industry that lasted for several months after the earthquake. But beyond the automobile industry, the global economy was not much affected, and even the automobile industry eventually recovered. Based on the criterion of "bend but not break," the global economy proved incredibly resilient to this shock.

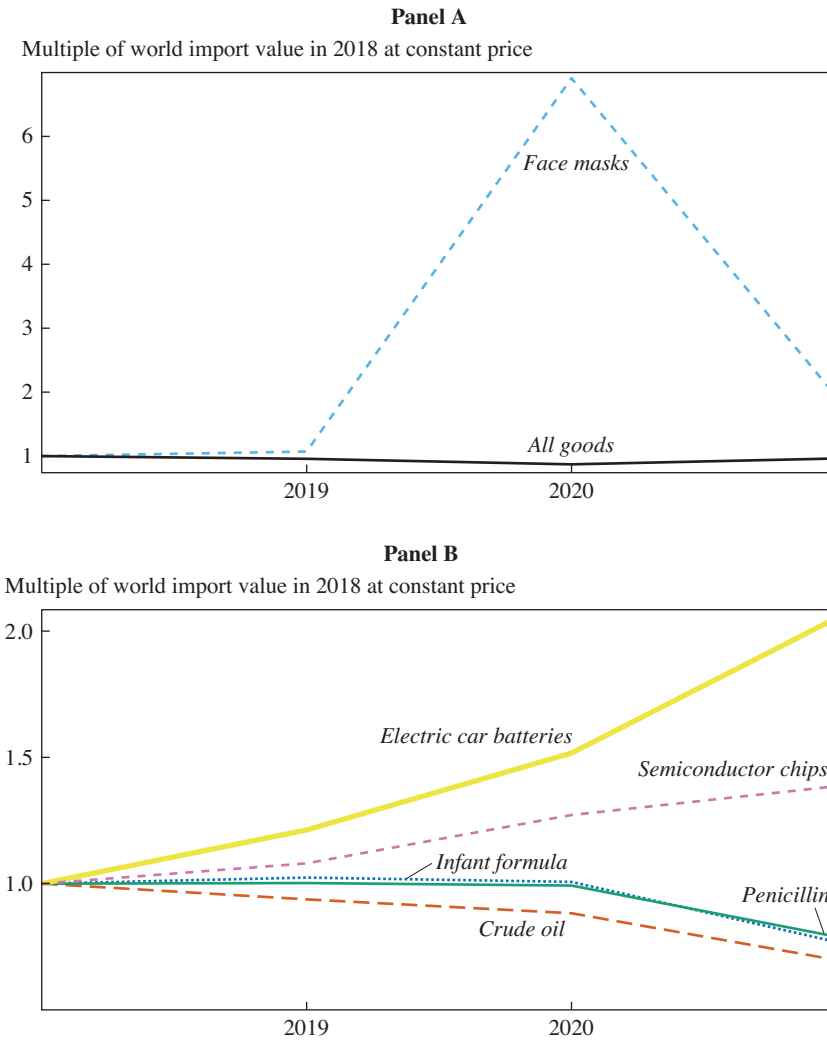
Returning to the COVID-19 pandemic, it is plausibly the largest shock the global economy has faced since World War II. It represented both a demand and a supply shock, as unemployment rose and income dropped sharply during lockdowns, while at the same time, production and commerce ceased in many countries. It affected all sectors of the economy, but some more than others (e.g., hospitality services versus finance or technology). It affected most countries on the globe, but not at the same time, as the infection waves were not synchronized across countries in different continents. In some sectors, such as food and medical products, delivery delays and shortages are critical (i.e., life threatening), while in others, such as autos or semiconductor chips, they are simply inconvenient.

How would one judge the resilience of GVCs and the global economy against this background? While, as noted earlier, we lack an agreed-upon benchmark against which resilience can be measured, we can look at how global trade, both in aggregate terms and in some key products, evolved during the pandemic years, keeping in mind that the pandemic is not completely over yet.

Figure 4 shows world trade flows since 2018 up to 2021, when the most recent COMTRADE data are available. Values are reported as a multiple of the 2018 world import value in constant prices. There was a negligible drop in all goods trade during the pandemic in 2020, and a full recovery in 2021, as seen already in figure 1. Imports of face masks, shown in panel A of figure 4, exhibit an interesting pattern: these imports increased sharply in 2020 reflecting the increasing use of face masks from China and Korea to meet the surge in domestic demand. Not only did international trade not let down the economy in this case, without imports domestic demand for face masks would not have been met by domestic producers. Growth in use of imports to supply electric car batteries and semiconductor chips grew in real terms during the pandemic, with imports of batteries accelerating. Use of imports to supply penicillin, infant formula, and crude oil has declined slightly as of 2021, but supply is not broken.

Khanna, Morales, and Pandalai-Nayar (2022) highlight two additional measures of resilience that can be measured in firm-to-firm transactions: (1) whether it is easy for firms to find new suppliers; and (2) whether firms

Figure 4. During COVID-19 Import Usage Was Bent Not Broken



Sources: COMTRADE and Bureau of Labor Statistics.

Note: World imports of goods are identified with six-digit Harmonized System (HS) codes: face masks (630790), penicillin not packaged for retail (300410), infant formula (190110), crude oil (270900), electric car batteries (850760), and semiconductor chips (854231). Nominal values from COMTRADE are converted to 2018 prices using import price indexes from the Bureau of Labor Statistics. Some indexes match nominal values exactly at the four-digit HS code: crude oil (EIUIP2709), penicillin (EIUIP3004), and semiconductor chips (EIUIP8542). Where such matches are not available in 2018, goods are matched to price indexes at higher levels of aggregation: face masks are matched to instruments and appliances used in medical, surgical, dental, or veterinary sciences (EIUIP9018); electric car batteries are matched to vehicles other than railway or tramway rolling stock, and parts and accessories thereof (EIUIP87); and infant formula is matched to food, beverage, and tobacco preparations (EIUIPIV). Values for all goods are deflated by the US GDP deflator as in figure 1.

maintain links with suppliers. The first measure is summarized, for each firm j and time period t by:

$$Entry\ Rate_{j,t} = \frac{N\ of\ supp.\ to\ j\ in\ t, who\ don't\ supp.\ in\ t-1}{\left[\left(N\ of\ supp.\ to\ j\ in\ t-1\right)+\left(N\ of\ supp.\ to\ j\ in\ t\right)\right]/2},$$

which measures the percentage of suppliers who are new in the current period. A market is resilient if the entry rate does not decrease after a shock. The second measure is summarized by:

$$Separation\ Rate_{j,t} = \frac{N\ of\ supp.\ to\ j\ in\ t-1, who\ don't\ supp.\ in\ t}{\left[\left(N\ of\ supp.\ to\ j\ in\ t-1\right)+\left(N\ of\ supp.\ to\ j\ in\ t\right)\right]/2},$$

which measures the percentage of suppliers from the last period that no longer supply in the current period. A market is resilient if the separation rate does not increase after a shock. A net separation rate may also be measured as:

$$Net\ Separation\ Rate_{j,t} = Separation\ Rate_{j,t} - Entry\ Rate_{j,t},$$

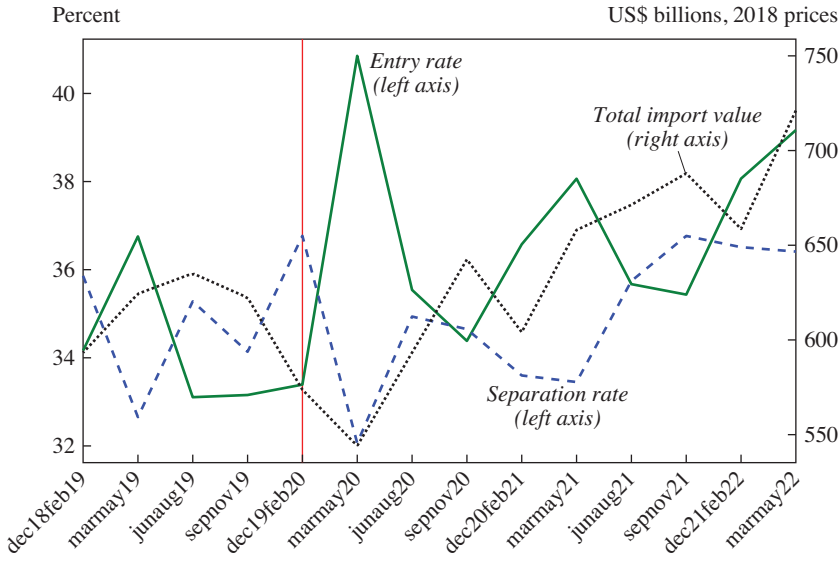
To evaluate the resilience of US supply chains during COVID-19, we construct these measures for US imports. The data source is Panjiva, which compiles the manifests for all container shipments coming into US ports.[18] Though the data do not include trade values, firm identifiers allow us to measure the extensive margin of trade. For each shipment, the data identify the US firm receiving the shipment (consignee) and the foreign firm sending the shipment (shipper).

Figure 5 shows the quarterly averages of the entry and separation rates across all importing firms in the United States during COVID-19 and a year before. Both measures have an average of about 35 percent, as many firms do not receive shipments from the same firms each quarter and receive shipments only once or twice a year.

By these measures, supply chains appear resilient in the quarter of March 2020 to May 2020, the first wave of the pandemic. In that quarter the total value of imports dropped around 13 percent in real terms relative to the same quarter in the previous year, reflecting reduced demand and prices as the country entered lockdown and, potentially, supply reductions

18. S&P Global, "Panjiva Supply Chain Intelligence," https://panjiva.com/.

Figure 5. US Firms' Relationships with Foreign Suppliers during COVID-19



Sources: Panjiva; US Census Bureau; and Bureau of Labor Statistics.

Note: The vertical line indicates the quarter before the COVID-19 pandemic begins. The entry rate measures the percentage of suppliers that are new in the current period, and the separation rate measures the percentage of suppliers from the last period that no longer supply in the current period. Total import value is the value of goods imports (not seasonally adjusted) reported by the US Census Bureau deflated by the import price index for all commodities.

elsewhere.[19] Despite this, the entry rate spiked to 42 percent. The separation rate also fell, as firms held on to previous suppliers. Since entry grew by more than separations, the net separation rate fell. It seems that during the pandemic domestic firms held on to existing relationships with trading partners and even pursued new ones, showing resilience in the context of declining total trade.

An explanation for these patterns is that COVID-19 was a shock to import supply and demand. On the one hand, international suppliers faced lockdowns and curtailed supply. On the other, there was a large reallocation of demand across products, as households scaled purchases of durables in

19. Total import values reported by the US Census Bureau, "Trade in Goods with World, Not Seasonally Adjusted," https://www.census.gov/foreign-trade/balance/c0015.html. In the Panjiva data the volume of container shipments also declines in March to May 2020 relative to March to May 2019.

anticipation of prolonged lockdown and also purchases of health-related goods, like face masks.[20]

To isolate the response of firms to only the supply-side shock, we exploit the time-varying nature of COVID-19 lockdowns across countries. Define the supply-side lockdown exposure of firm *j* in time *t* as:

$$\textit{Lockdown Exposure}_{j,t} = \sum_i s_{i,j,t_0} \times \left(\textit{Lockdown stringency in country } i \textit{ at time } t\right),$$

where s_{i,j,t_0} is the share of firm *j*'s import volume (measured in twenty-foot equivalent container units) from country *i* in t_0, the quarter December 2019 to February 2020, and lockdown stringency in country *i* at time *t* is the Oxford University lockdown stringency index (Wade and others 2023).[21] Time intervals are three-month quarters, so the stringency index in each time period is the average for the quarter. Lockdown exposure is zero for all firms prior to March 2020 and then varies across firms and quarters after the pandemic begins.

To isolate responses to this supply side shock we estimate the regression:

$$\textit{Net Separation Rate} = \alpha_j + \alpha_t + \beta_1\left(\textit{Lockdown Exposure}_{j,t}\right) + \beta_2\left(\textit{Lockdown Exposure}_{j,t} \times W_j\right) + \epsilon_{j,t},$$

where α_j is a firm fixed effect and α_t is a quarter fixed effect. The quarter fixed effects control for changes in demand in the US economy during the pandemic. W_j stands for alternative firm characteristics hypothesized to shape resilience, namely, the share of import volume (again in twenty-foot equivalent container units) of goods that are differentiated rather than sold on an organized exchange or with reference prices, according to Rauch (1999); the number of suppliers to the firm at t_0; and the total volume of imports at t_0.

Results are reported in table 1. Across all columns, and in contrast to the aggregate pattern in figure 5, supply-side lockdown exposure increases

20. Consistent with this, when replicating figure 5 using only firms importing either face masks or electric car batteries, the spike in entry rates is somewhat larger in the quarter starting March 2020.

21. Our World in Data, "COVID-19: Stringency Index," https://ourworldindata.org/covid-stringency-index.

Table 1. Resilience of US Importers to Supply-Side Shocks

	(1)	*(2)*	*(3)*
Variable	*Net sep. rate*	*Net sep. rate*	*Net sep. rate*
Lockdown exposure [0, 1]	7.14***	5.71***	5.60***
	(0.52)	(0.43)	(0.43)
Lockdown exposure × share differentiated in t_0	−2.02***		
	(0.35)		
Lockdown exposure × Z (number of suppliers in t_0)		3.13***	
		(0.30)	
Lockdown exposure × Z (import volume in t_0)			1.73***
			(0.63)
Number of observations	913,822	913,822	913,822
R^2	0.04	0.04	0.04
Firm fixed effects	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes

Source: Authors' calculations using separation rate data from Panjiva and lockdown exposure from Oxford University.

Note: Number of suppliers in t_0 and import volume in t_0 are normalized as Z-scores by subtracting off the sample mean and dividing by the standard deviation. Standard errors in parentheses are clustered at the firm level.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

net separations. This confirms that the aggregate pattern of decreasing net separations is explained by demand factors.

There is heterogeneity in responses across firms. Perhaps surprisingly, firms whose imports are more differentiated by the measure of Rauch (1999) have fewer net separations as they maintain relationships even in the presence of a supply shock. In contrast, firms with more suppliers and those with larger overall import volumes experience greater net separations, potentially because they have a greater pool of relationships on which to draw. Khanna, Morales, and Pandalai-Nayar (2022) find similar patterns of heterogeneity when studying the response of firms to supply-side shocks generated by lockdowns in Indian districts. Using firm-to-firm transactions data, similar to those in the Panjiva data but covering also domestic transactions, they exploit the differential exposure of firms in that state to supplier risk during the pandemic. Because the severity of lockdowns differed greatly across Indian states, suppliers located in states with more strict lockdowns experienced a larger disruption. In an event study, they show that firms with suppliers exposed to more-severe lockdowns experienced an increase in net separation rates, a similar result to column 1 of table 1. The intensive margin of trade as measured by the value of inputs and firm output also suffered. They also document that these disruptions were less

severe for supply chains trading more complex products.[22] Though the paper does not attempt to nail down the mechanism behind this somewhat unexpected finding, a plausible explanation is that firms that produce more-complex products value supplier-specific relationships more and take great care to preserve them when faced with uncertainty and disruption.

The importance of relationship-specific investments in global supply chains, and international trade more generally, has been emphasized in the literature.[23] Seen through this lens, it is perhaps not surprising that the US firms that appear in the Panjiva data and trade internationally did not on average sever existing relationships during the pandemic. Furthermore, the additional demand for certain products, for example, face masks or electric batteries, may have led them to seek new trading partners. At any rate, though supply-side shortages due to lockdowns contributed to net separations, overall the extensive margin of trade resulting from the combination of supply and demand forces indicates—contrary to popular claims—strong resilience during the pandemic, as shown in figure 5.

This message is reenforced in a few other papers that explore the economic effects of the pandemic. Stumpner (2022) studies regional lockdowns in China and shows that with the exception of the Shanghai lockdown of April 2022, the other regional disruptions in China due to COVID-19 had no effect on international trade. He concludes that regional lockdowns could still reduce aggregate trade, but only if they were exceptionally stringent and implemented in an economically important province such as Shanghai.

Using data for sixty-four countries, Bonadio and others (2021) document a large contraction of economic activity (an average 29.6 percent drop of GDP) during the pandemic. But only a small fraction of this contraction (about 23 percent) can be attributed to foreign shocks that are intermediated through final and intermediate goods trade. Moreover, simulations based on a global network model indicate that global supply chains alleviated the pandemic-induced contraction: the downturn in the sixty-four countries of their sample would have been worse on average if there were no international trade and supply chains relied on domestic inputs only. The reason is that elimination of foreign inputs would have increased reliance on domestic inputs, which were also affected by the pandemic.

Lastly, looking at world- and country-specific GDP rather than trade reinforces the message of resilience. Clearly, the world economy took a major

22. They use a few alternative measures of complexity: the number of products a particular firm buys from suppliers; the share of total purchases accounted for by differentiated products; and the average number of inputs that are necessary to produce each firm's product.

23. See World Bank (2020) for an overview.

hit with the COVID-19 shock in 2020. But most countries rebounded. World GDP contracted in real terms by –3.2 percent in 2020, but increased 5.9 percent in 2021. The US economy grew on par with global growth in 2021, and the recovery was even faster in several regions according to the World Bank (2023).[24] Even though the pandemic is still not over and many of its effects will be felt for many years to come, the world economy is far from broken. Global growth in 2022 is estimated at 2.9 percent. This is indeed a slowdown relative to the rapid recovery of 2021, but still greater than the 2019 rate of 2.5 percent (World Bank 2021). So, once again, based on the criterion "bend but not break," the world economy proved resilient during the pandemic.

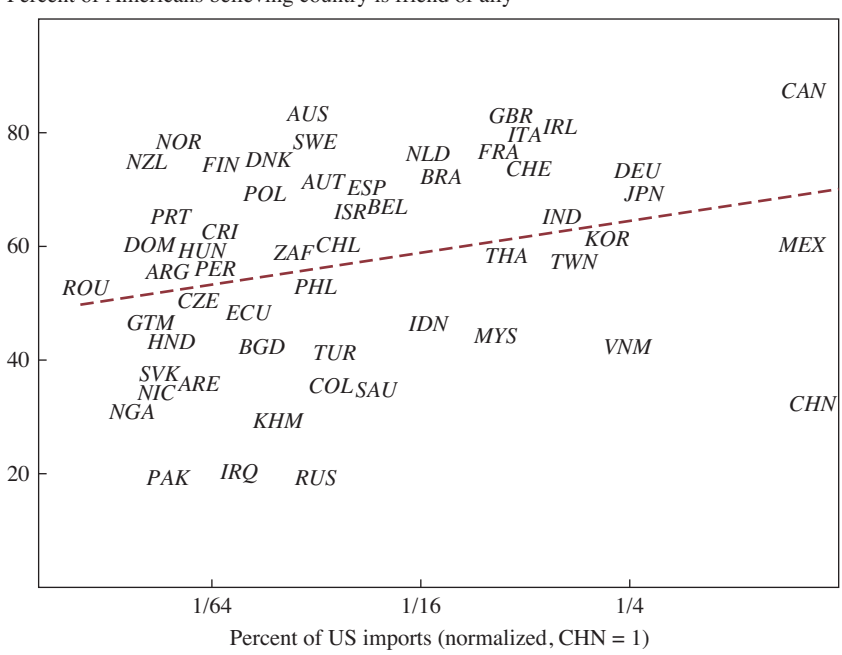
In sum, despite the prominence of resilience concerns in the public debate in the past three years, the evidence to date provides no support either for the view that global supply chains were not resilient during the pandemic or that the world economy would have been more resilient if there had been less dependence on foreign inputs and trade.

INVASION OF UKRAINE AND ENERGY CRISIS The alleged effects of trade as well as COVID-19 on advanced economies' labor markets and inequality supplied many arguments against free trade and globalization, but no single one of them seemed to cause a reversal of attitudes and long-term trends. Nevertheless, these developments might have prepared the ground for what has turned out to be a game changer in the debate and policy on globalization: the quest for resilience, this time not to economic shocks or natural disasters, but to geopolitical turmoil and the emergence of national security as a major reason for rethinking globalism. This is turn fed new demands for reshoring as well as for a rather new concept in trade policy: friendshoring, that is, trading only with, or predominantly with, friendly nations.

The catalyst for this new phase in the deglobalization movement was the invasion of Ukraine by Russia in February 2022, which was followed by a major energy crisis. At this point, the vulnerability of Europe created by relying on a single source (i.e., Russia) for a large share of its energy imports became apparent. The evidence on supply chain resilience during COVID-19 provided by Khanna, Morales, and Pandalai-Nayar (2022) for Indian firms as well as by our results for US firms suggests that the private sector had taken steps to cope with exogenous shocks, at least in those cases where disruptions would be expected to have severe economic

24. In 2021, real GDP growth in the United States was 5.9 percent, but 7.9 percent in South Asia, 7.2 percent in East Asia and the Pacific, and 6.7 percent in Eastern Europe and Central Asia.

Figure 6. America Is Already Friendshoring



Sources: YouGov (2017) and US Census Bureau.

Note: Countries are classified as unfriendly if less than 50 percent of Americans believe the country is a friend or ally. Figure includes only countries with greater than 0.1 of US imports in 2022. The slope of the dashed line is 4.04 (SE = 1.87).

implications due to relation-specific investments. But the experience with Russia revealed that at least Europe was ill-prepared to deal with disruptions due to major geopolitical shocks. The concerns about Europe and Russia have led—by extrapolation—to more general concerns about the resilience of global supply chains to geopolitical risks and to demands to fundamentally reorganize international relationships, so as to decouple from any country that is perceived as unfriendly by the United States. China tops the list.

A natural question in this context is whether, beyond specific instances of government-led sanctions, market forces have led countries already to orient themselves toward trading with friends. In figure 6 we evaluate whether the United States is in fact friendshoring by relating the import share of a country in 2022 to the percentage of Americans believing the

country is a friend or ally in the most recent poll by YouGov (2017). For many trading partners, a majority of Americans believe the country is a friend, and there is a positive and statistically significant association between this variable and a country's share of imports. China, the largest trading partner, is in fact an outlier, with just 26 percent of Americans believing it is friendly and 6 percent believing it is an ally.

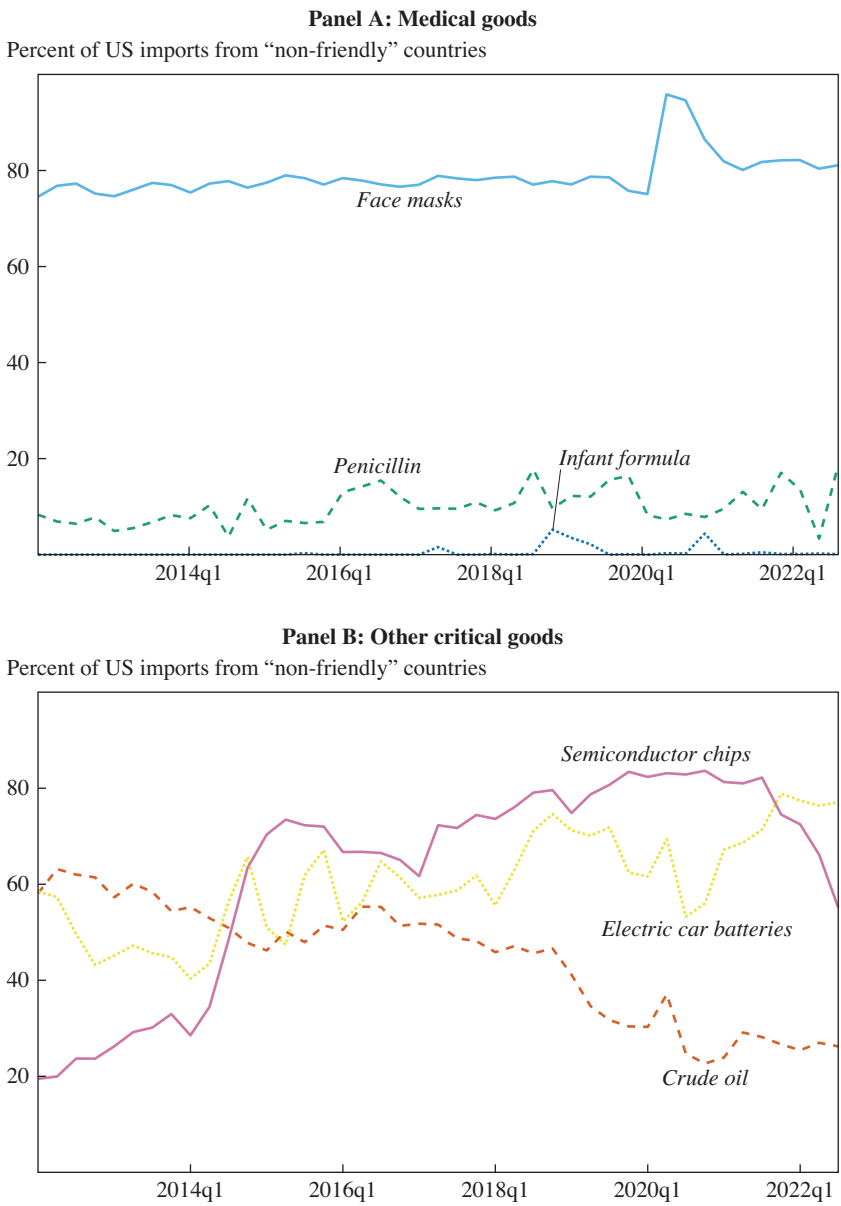
The survey data reveal the limits of friendship as an organizing principle for trade. For many countries, Americans participating in the YouGov survey respond that they are unsure whether the country is an ally, friend, unfriendly, or an enemy. Indonesia, Malaysia, and Vietnam are all important sources of imports, including semiconductors, but less than 50 percent of Americans believe they are a friend or ally. There is a pro-European bias to surety, with, for instance, Japan ranking below Germany despite nearly identical import shares.

Though the aggregate import data suggest an already high degree of friendshoring in the United States, it is possible that high dependence on non-friendly nations for the import of some critical products makes the United States vulnerable to geopolitical risks (as was the case with Europe and energy import from Russia, for example). For this reason, we focus on some high-profile products next. These products are considered critical either for health and nutrition reasons or because they represent important inputs in the production processes of a modern economy. Figure 7 reports, for health (panel A) and other strategic goods (panel B), the share of US imports coming from countries where less than 50 percent of Americans believe the country is a friend or ally, or non-friendly countries, measured in 2017. We hold the classification of countries constant to illustrate dynamics in import shares. There are several observations.

First, in the category of health products, face masks, of which 73 percent came from China in 2022, is an outlier in terms of reliance on non-friendly countries. Non-friendly countries provide less than 20 percent of penicillin imports and less than 5 percent of infant formula imports. Evenett (2020) makes a similar point in his study of pre-pandemic international sourcing patterns of medical goods and medicines for France, Germany, the United Kingdom, and the United States.

Second, as noted earlier, in the case of face masks, imports from China provided important relief and increased resilience during the pandemic. Note the spike of imports from non-friendly nations during the second quarter of 2020 in figure 7. Hence, decoupling from unfriendly nations may increase resilience to specific geopolitical shocks but decrease resilience to

Figure 7. Most US Imports of Several Strategic Goods Come from Non-friendly Countries



Sources: YouGov (2017) and US Census Bureau.

Note: Countries are classified as non-friendly if less than 50 percent of Americans believe the country is a friend or ally. Imports are identified with six-digit Harmonized System (HS) codes: face masks (630790), penicillin not packaged for retail (300410), infant formula (190110), crude oil (270900), electric car batteries (850760), and semiconductor chips (854231).

other shocks unrelated to politics. The general point is that it is impossible to think about resilience without specific reference to the type of shock to which resilience is sought.

Third, panel B of figure 7 shows how import dependence on non-friendly nations has evolved in the past decade for some strategic goods. The sharp increase in the case of semiconductor chips reflects the rising popularity of the foundry model in the past two decades, with many foundries located in Malaysia, Vietnam, and China (see table A1 in the online appendix)—countries that are perceived as non-friendly by a majority of Americans in the YouGov survey. Similarly, in the case of electric batteries, the growing dependence on China is evident in the figure. However, import shares for some goods appear responsive to policy. The share of crude oil sourced from non-friendly countries has fallen from 60 percent to about 20 percent, as the shale boom made the United States a petroleum exporter. Although in 2020 about 80 percent of semiconductor chips were imported from non-friendly countries, in 2022, during which the United States articulated a friendshoring policy with regard to the sector, this share had fallen dramatically to about 55 percent. This reflects growth in the import share of Taiwan from 5 percent to 10 percent; Ireland, from 2 percent to 8 percent; and Israel, from 1 percent to 4 percent.

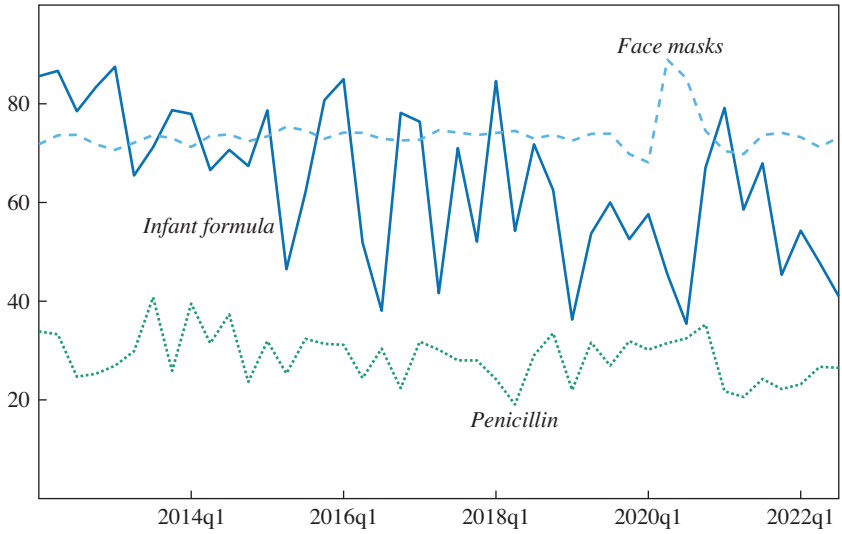
Dependence on non-friendly countries is only one among several possible factors affecting supply chain resilience. High concentration in the market of a specific product will in general imply high dependence of buyers on suppliers, making this market less resilient to supplier-side risk—be it risk associated with exogenous shocks (e.g., natural disasters) or risk arising from the exercise of supplier market power. Given that friendship is a volatile concept (ironically, two of the countries viewed as friends of the United States today, Germany and Japan, were its main enemies during World War II), resilience in its broadest form implies availability of multiple alternatives for sourcing a product and low concentration.

Figure 8 and table A1 in the online appendix show how concentrated the markets for the above strategic product imports are in the United States. Concentration can signal a lack of resilience because it implies fewer alternatives to choose from if one supplier fails or if demand increases unexpectedly. Figure 8 reports the percentage of US imports from the country with the largest import share in each product category, and table A1 shows the import shares of the five largest importers in each case. By these measures, some goods markets appear less resilient than when viewed through the lens of supply by friendly countries. Approximately 75 percent of infant formula comes from two countries, Ireland (45 percent) and Mexico

Figure 8. Markets for Imports Are Highly Concentrated

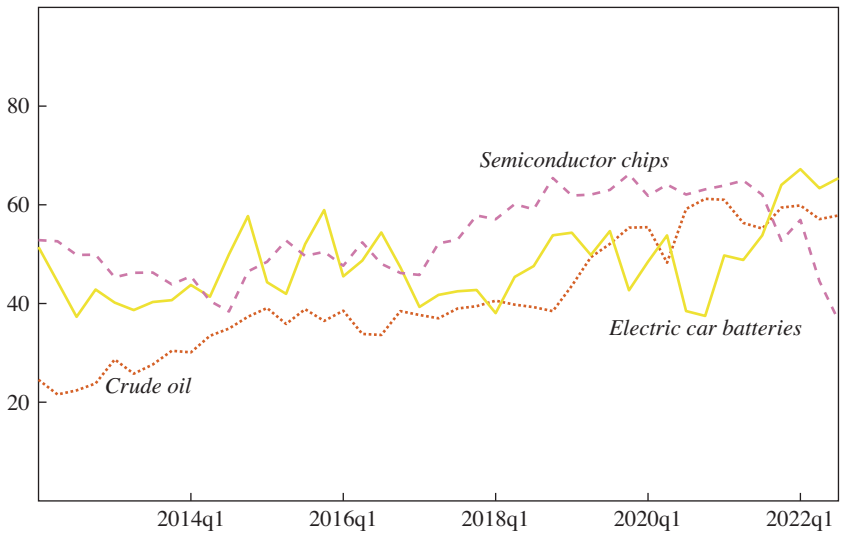
Panel A: Medical goods

Percent of US imports from country with the largest import share



Panel B: Other critical goods

Percent of US imports from country with the largest import share



Source: US Census Bureau.

Note: Imports are identified with six-digit Harmonized System (HS) codes: face masks (630790), penicillin not packaged for retail (300410), infant formula (190110), crude oil (270900), electric car batteries (850760), and semiconductor chips (854231).

(30 percent). This concentration is due to differences in market access: Ireland and Mexico have duty-free access through free trade agreements, whereas most other countries are subject to a tariff of 14.9 percent to 17.5 percent depending on the content of the formula (Casey 2022). Stringent labeling and food safety requirements, which are not harmonized with those required by regulators in Europe, further restrict imports (Casey 2022). Similarly, in the case of crude oil, about 70 percent of US imports in 2022 came from two (friendly) countries, Canada and Mexico. The main takeaway from these data is that key product markets in US imports are highly concentrated. Given that the main suppliers in most of these markets are friendly countries, this state of affairs does not imply vulnerability to geopolitical risk—at present! But it does imply vulnerability to other country-specific risks as well as to a potential change in international relations in the future.

Measures of import market concentration are of course imperfect proxies of resilience. Ideally, one would like to know elasticities of substitution and export supply and import demand elasticities at a highly disaggregate level to assess a product market's resilience to a shock. For instance, in the case of a demand shock (say, an increase in the import demand for face masks due to COVID-19), a market would be characterized as resilient if the export supply of face masks by the rest of the world were highly elastic. And similarly, in the case of a supply shock (say, a decrease in supply of inputs due to a lockdown), a market would be considered as resilient if the import demand were highly elastic. Export supply and import demand elasticities are notoriously difficult to estimate credibly. Fajgelbaum and others (2020) use the tariff variation induced by the recent trade war between the United States and China to estimate these elasticities at the HS-10 level of aggregation. Interestingly, they cannot reject the hypothesis that the export supply curve facing the United States is infinitely elastic. Similarly, they cannot reject that the export supply curve of the United States facing foreign trade partners is infinitely elastic. Taken at face value, the elasticities imply high resilience to demand-side shocks.

However, these elasticities are measured at a level of aggregation that may still be too high for judging a sector's resilience. The vulnerabilities that are often cause for concern play out at a much more disaggregate level. Consider semiconductors for instance. According to table A1 in the online appendix, Taiwan has only around a 10 percent share in US semiconductor chip imports. However, as a joint report by the Boston Consulting Group and Semiconductor Industry Association shows, Taiwan dominates the market for logic chips in the advanced nodes (10 nanometers or below),

which are required for compute-intensive devices and smartphones (Varas and others 2021). If the production technology in these products follows Leontief, then there will be no substitutes for these highly specialized chips made in Taiwan. Industry reports suggest that this is indeed the case. But obtaining credible estimates of substitution elasticities at that level of disaggregation has been elusive so far. Future micro-oriented research might be able to make progress on this front.

Along the same lines, our discussion so far has focused on resilience as it relates to imports—given that this paper is about deglobalization and the forces that are shaping it. But resilience of supply chains depends as much on domestic market structure and competition as on international factors. For example, an industry that sources its inputs from a single domestic supplier might be plausibly sheltered from geopolitical risk but would have little resilience to other shocks. A complete assessment of resilience therefore requires careful modeling of the entire industry under consideration, taking into account both domestic and international firm-to-firm relationships. This could be another area for policy-relevant micro research in the future.

In general, the evidence to date does not suggest any strong dependence of the United States on imports from non-friendly countries—on average—though there is dependence is some critical sectors. However, even in these sectors, the recent experience with COVID-19 suggests that countries regarded as non-friendly alleviated rather than caused critical bottlenecks. In addition, it seems that the private sector, even without government intervention, has been slowly decoupling from non-friendly countries in recent years, especially in semiconductors and crude oil. However, imports of critical products are highly concentrated among a few countries. Independent of geopolitical risk, this is a cause for concern as it suggests potential vulnerabilities to idiosyncratic supplier-side risk.

DUAL-USE GOODS AND CONCERNS ABOUT NATIONAL SECURITY The national security argument is not new. It was initially deployed by the Trump administration to impose tariffs on aluminum and steel, and subsequently endorsed by the Biden administration, which declared these tariffs were nonnegotiable. As discussed earlier, it has also been invoked to justify policies encouraging decoupling from non-friendly nations in the imports of critical products in order to make their supply resilient to geopolitical risk. But it has reached new significance since the last quarter of 2022 when it was employed proactively to stop non-friendly countries, China specifically, from developing military capabilities by exploiting trade with the United States to their advantage. The cornerstone of this argument is the increasing importance

of "dual-use goods," that is, goods that have both civilian and military uses. A complete list of such products can be found in Singapore, where under the Strategic Goods (Control) Act permits are required for transshipment of dual-use goods.[25] The list of dual-use goods is vast, including aluminum and steel alloys above a certain tensile strength, semiconductor chips, and machine tools.

The main category targeted to date is the semiconductor sector. Semiconductors are an integral component of various consumer products, such as cars and smartphones, but are also used in dual-use goods, such as civilian and military aircraft. Moreover, they are used in supercomputing and artificial intelligence, fields that have potential national security implications. Motivated by these considerations, the United States in October 2022 announced sweeping export controls in the semiconductor industry targeting China. The United States does not export many semiconductor products directly to China. However, the export controls targeted third-country chip manufacturers who use US software or US machines in their manufacturing facilities. According to the restrictions, any semiconductor made with US technology for use in supercomputing or artificial intelligence can be sold to China only with an export license issued by the United States, which will be difficult to obtain. Given that almost every semiconductor is produced using US technology, this rule effectively covers the entire global industry. Third-country producers had no say in the matter. They have two choices at this point: either obtain the required export licenses by the United States or cease to use US technology and equipment. Hence, the use of the phrase "weaponized interdependence" to characterize how the United States has used the interdependence inherent in trade and global supply chains to force its trade partners to go along with its economic war against China. In addition, the United States barred its citizens from working with Chinese chip producers except with specific approval. With these measures the United States is seeking to prevent China from advancing technologically in sectors that are crucial to national security.

Interestingly, none of these restrictions were the result of lobbying efforts by the domestic or global semiconductor industry (in fact, the semiconductor industry has become so globally interconnected that it is one of the strongest advocates of free trade). They were decided top-down by the US government based on concerns about the increasing militarization of China and military applications of dual-use goods. The global semiconductor industry

25. Singapore Customs, "List of Dual-Use Goods," https://www.customs.gov.sg/businesses/strategic-goods-control/strategic-goods-control-list/list-of-dual-use-goods/.

was asked to adjust or lose access to US technology. To the extent that these new policies contribute to deglobalization today, it is fair to say that deglobalization is not market-driven; rather, it is the result of government-led actions and policies that had little support, at least initially, from the private sector.

Taken at face value, national security is the most powerful argument against unconstrained, market-driven globalization to date. It is also the most problematic as it cannot be verified directly by researchers, market analysts, or journalists—one has to place faith in the government's intelligence sources. Nevertheless, the argument has found bipartisan support in the United States. Furthermore, a broad interpretation of "dual use" might lead to many goods facing restrictions—including clothing or medicines used by the military. Accordingly, it has the potential to lead to broad, sweeping restrictions in several sectors and a major economic war if China retaliates. An alternative interpretation of the recent export restrictions in semiconductors is that they have little to do with national security but aim instead to contain China's economic development, as the United States has realized that industrial policy alone will not be sufficient to outcompete China in the future.[26] If this is the case, the new restrictions mark the end of an era of globalism and economic cooperation and the onset of another cold war.

III. Consequences

We conclude with some thoughts about the potential consequences of gradual deglobalization. Given that, as shown in section I, deglobalization is not showing in the data yet but is nevertheless plausible given recently adopted policies, any statements about its likely effects in the future are highly speculative.

In the short and medium run, one would not expect any dramatic effects as the world economy is slowly transitioning to a new state. One should also distinguish between level and direction of change. The level of globalization remains extremely high by historical standards, though the direction of change is toward deglobalization. As global supply chains are reorganized

26. See Sullivan (2022): "On export controls, we have to revisit the longstanding premise of maintaining 'relative' advantages over competitors in certain key technologies. We previously maintained a 'sliding scale' approach that said we need to stay only a couple of generations ahead. That is not the strategic environment we are in today. Given the foundational nature of certain technologies, such as advanced logic and memory chips, we must maintain as large of a lead as possible" (pars. 43–45).

across the world, a new international economic system may emerge, one that relies heavily on bilateral and regional agreements as well as partnerships among friends. This process may create opportunities for countries that are well positioned to take advantage of this new economic environment.[27] But even though trade will likely survive new geopolitical tensions, the consequences of the newly emerging economic system on the global economy may turn out to be more severe in the long run.

III.A. Resilience

Given that many of the recent policy restrictions were motivated by the quest for resilience, be it resilience to natural and economic shocks or resilience to geopolitical risk, it is natural to ask whether the new system envisioned by policymakers would make the economy more resilient. As emphasized throughout this paper, this question is impossible to answer without a well-defined metric of resilience. Developing such a metric requires value judgments by our society; it is not simply an economic question. For instance, ongoing work by Carvalho, Elliott, and Spray (n.d.) makes progress on this conceptual challenge by defining a closely related concept: that of a firm being "essential" in a case where "demand for key goods at current prices cannot be met without that firm producing . . . i.e., there is no way of re-routing demand through the supply network such that other firms can take up the slack" (1, 4). This is a useful starting point, but it still leaves open the question of what "key goods" are. From the point of view of some US consumers, a new car of the latest model year might be a key good. Similarly, if one focuses on the total output of the economy, the question becomes how large a decline to a major shock (such as COVID-19) is a society willing to tolerate.

These qualifications notwithstanding, reviewing the theoretical arguments and evidence to date suggests that future resilience—no matter how exactly this is defined—could go either way.

On the theory side, recent work by Elliott, Golub, and Leduc (2022) shows how modern production networks create the potential for critical bottlenecks. Private optimization is not sufficient to avoid severe disruptions, so there is scope for intervention by a social planner. In this context, reshoring in critical sectors is potentially justified.

On the empirical side, the evidence is mixed. As discussed earlier, existing supply chains in almost all sectors proved resilient to the pandemic.

27. See, for example, the response to the US-China trade war documented in Fajgelbaum and others (2022).

But energy supply in Europe turned out to be highly vulnerable during the Ukraine war. In general, as emphasized earlier, resilience cannot be judged without reference to specific shocks. Caselli and others (2020) make this point forcefully when considering the question of whether international trade makes a country more or less resilient (i.e., in the sense of being exposed to volatility) to shocks. As they point out, the answer depends on whether the shocks are sector-specific or countrywide. A common view is that trade makes countries more vulnerable to shocks, as trade encourages specialization that implies high exposure to sector-specific shocks. However, when countrywide shocks are important, international trade may make a country more resilient by reducing its exposure to domestic shocks. Using a quantitative model of trade, they find that in recent decades the second type of shock dominates, so that international trade has reduced economic volatility for most countries. The evidence reviewed in section II pertaining to COVID-19, a shock that was not just countrywide but global—however, not synchronized across countries—reinforces this view.

In short, unless a sector is highly dependent on a single import source (as is the case with the dependence of the energy sector in Europe on Russia), international trade seems to contribute to resilience, not compromise it. Hence, it is unlikely that trade restrictions will improve countries' resilience.

III.B. Efficiency, Innovation, and Long-Term Growth

It may be hard to clearly identify the causal effects of global supply chains and openness on efficiency and growth through econometric studies, but the consensus is that many of the features that characterize modern trade, such as hyper-specialization, firm-to-firm relationships, and technology and knowledge transfer across firms and countries, played a crucial role in promoting growth and technological progress over the past three decades. The stagnation, if not reversal, of the open policies of the past by the United States, which more than anyone else had embraced them, naturally raises questions about the future of growth and innovation under the new regime. The recent export restrictions in semiconductors represent a major blow to the technological advancement of China (as intended), at least in the short run. How China will cope remains to be seen. The hope of the United States is that the same policies aimed at containing China, combined with well-thought-out industrial policy, could spur a new wave of growth and innovation in the United States. But industrial policy has a mixed record, and it remains to be seen whether its recent incarnation will prove beneficial to the US economy in the long run. Further, a slowdown in Chinese growth and innovation could slow the United States and global economy.

Although the future is highly uncertain in this changing landscape, there are reasons to be concerned about future growth prospects. In a study completed in 2021, well before the sweeping export restrictions targeting China were put in place, the US Chamber of Commerce issued a report studying the aggregate costs of a potential US-China decoupling as well as its industry impacts (US Chamber of Commerce 2021). The report concluded that the costs would be uncomfortably high; for instance, abstracting from other measures, a 25 percent tariff applied to all two-way trade would imply an annual GDP loss for the United States on the order of $190 billion by 2025. The aggregate effects would be orders of magnitude larger if one considered restrictions on investment, people, and the flow of ideas. The impacts on specific industries such as aviation, semiconductors, chemicals, and medical devices would also be substantial over the next decade.[28]

Along the same lines, a recent paper by Thun and others (2022) introduces a new concept, "massive modularity," that characterizes many production processes today; they argue that the presence of massive modularity makes it extremely hard to decouple, reshore, and generally reorganize economic activity across borders. Massive modular systems involve several modules that are interconnected with each other, can experience innovation independent of each other, and can be broken into smaller, more specialized modules, each of which can again experience independent innovation. Different firms, located in different countries, specialize in different modules, making production structures extremely complex. For example, the CEO of Pfizer once stated that the company's COVID-19 vaccine "requires 280 different materials and components that are sourced from 19 countries around the world" (Breuninger 2021, par. 5). The vast complexity of modern production poses a challenge for policy as measures aimed at reducing risk or promoting domestic industries may have unintended consequences. In general, rebuilding massively modular industries in all their complexity on a national level is a Herculean task. Even if it doesn't fail, it will certainly take many years to accomplish. Given that the sectors characterized by this high complexity are precisely those sectors that are key to innovation and growth, this effort will likely slow down growth in the United States and global economy.

Decoupling between the United States and China in particular also threatens the pace of global innovation. Most models of long-term growth

28. In aggregate, IMF researchers estimate losses on the order of 8–12 percent in some economies once technological decoupling is considered (Aiyar and others 2023).

emphasize the role of population in research and development.[29] With over 1.4 billion people, China is expected to have a lot of new ideas and develop advantages. For example, China is a global leader in 5G communications technology and consistently files more artificial intelligence patents than any other country (Li, Tong, and Xiao 2021). Xie and Freeman (2019) attribute 37 percent of global citations to scientific articles written by Chinese researchers. Though citations are an imperfect measure of innovation, the scale is unquestionable. Moreover, a significant amount of this research has been done with American coauthors.[30] As political tensions have increased, this scientific collaboration has come under scrutiny. In 2018, the US Justice Department China Initiative began investigating US-based scientists under suspicion of intellectual property theft on behalf of the Chinese government, and the US National Institutes of Health began investigating hundreds of scientists for nondisclosure of research funds from China. Jia and others (2022) show that around this time, there was a marked decline in publications by and citations of scientists with previous collaborations with scientists in China, even if these scientists were not themselves subject to investigations. US-based scientists of Chinese heritage were hit hardest. In interviews, scientists cited new administrative oversight, including frequent consultation with their university's administration to navigate regulations about collaboration, and a feeling that they had to choose between access to US research dollars and their collaborations with scientists in China.

While US authorities have legitimate reasons to enforce intellectual property rights over commercialized technology developed at private firms, the type of pre-commercial basic research done at universities is often done without an expectation of patent. Indeed, many university scientists are not motivated by profit and recognize that their research is most beneficial when it disseminates quickly and can be built on by other scientists. As Chinese science continues to advance even under US export restrictions, barriers to interaction between US scientists and Chinese scientists could plausibly retard innovation in the United States.

III.C. Inflation

Inflation emerged as a major concern in 2021. No one claims that the recent increases in prices are due to trade restrictions. The US-China trade war may have increased the prices of targeted products but it had small

29. See, for example, Kremer (1993).

30. Along the same lines, US universities enrolled 129,440 Chinese graduate students in 2018–2019 (Feldgoise and Zwetsloot 2020).

effects on the US Consumer Price Index.[31] However, one of the main presumed benefits of open trade is its effects on lowering consumer prices. The low inflation the United States and many other advanced economies enjoyed in the past two decades, despite aggressive monetary policies, is in part due to globalization. International sourcing allowed firms to reduce their costs, and even though these cost reductions were not always passed through to consumers in the form of lower prices, so that many firms increased their profit margins, international competition kept price increases in check.[32] On the labor side, unions and workers in general had little bargaining power when their jobs could be outsourced to low-wage destinations (or to machines and robots). One may lament the effects that globalization and technology may have had on the American worker, but the flipside of these forces is that they kept prices low for consumers. If, in the new era, product and labor markets are shielded from foreign competition in the form of trade or immigration, the price of work and goods will rise. This is especially relevant in a period characterized by labor and other supply chain shortages.

III.D. Within-Country Inequality

On the other hand, the empowering of the American worker has been one of the stated objectives of the new stance toward globalization in the United States. The question then is, Will deglobalization halt and possibly reverse the rising inequality that emerged during the era of hyper-globalization? As the *Deaton Review* argues, inequality is a complex phenomenon with many dimensions—social, political, and economic—so that there is no simple answer to this question. However, the experience of the past two years should give one pause. In the United States, nominal median weekly earnings went up by 10 percent from January 2021 to December 2022. But due to 14 percent inflation, real wages actually declined by 4 percentage points.[33] Although factors other than deglobalization could be responsible for this trend, one thing that is certain is that the American worker was not better off in 2022 compared to past years.

31. Studies of the price effects of the trade war find complete pass-through of the tariff increases on import prices, though consumer prices were less affected. See Fajgelbaum and others (2020) and Cavallo and others (2021).

32. See De Loecker and others (2016) for a discussion of pass-through of cost reductions on prices.

33. Weekly earnings are median usual weekly real earnings for wage and salary workers employed full time, age 16 years and over, and inflation is the Consumer Price Index for all urban consumers, both reported by the Federal Reserve Bank of St. Louis.

III.E. Global Inequality

The effects of potential deglobalization on global inequality might be easier to determine. The past three decades saw a sharp reduction of global poverty, driven primarily by the growth of East Asia, and a decline in global inequality.[34] Though many factors contributed to these developments, the consensus among economists is that the opening of long-closed borders, the growth of trade between countries, and the establishment of the modern global trading system played an important role.

Such progress seems less likely in a future deglobalizing world where advanced countries are turning inward. Large developing countries, such as India, may still find a way forward if they implement appropriate policies by relying on their own large domestic markets and by taking advantage of the void left by the United States decoupling from China to advance their economic integration with the rest of the world. However, the lessons learned from trading with China will not be forgotten in advanced economies; India could expect to encounter a less friendly reception in the United States if its low-wage labor represented a threat to US labor markets or if it grew to the point that it threatened the technological dominance of the United States.

For smaller, low-income countries the prospects are much bleaker. Without access to lucrative foreign markets, there is no clear path for such economies toward growth, poverty reduction, and development, as Goldberg and Reed (2023) show. Increasing emphasis on environmental and labor standards as well as product regulation (such as sanitary and phytosanitary standards) raise entry barriers for poorer countries and may lead an increasing share of trade to be within the set of high-income countries that can comply, rather than between high-income and low-income countries.

Policies to fight climate change may further contribute to divergence between advanced and developing countries. If, for instance, countries sharing similar interests and characteristics (e.g., the prosperous economies of G7, or North America, or Europe) form climate clubs, then countries like India with high emission levels may find themselves excluded and potentially facing punitive tariffs.

III.F. Peace

A final, long-run consideration regards the effects of deglobalization on peace. Indeed, one of the strongest motivations for free trade has been the belief that it promotes peace and political stability. The predecessor of the

34. See Goldberg and Larson (2023) for a discussion of these trends.

European Economic Community, the European Coal and Steel Community, for instance, was designed with the explicit goal to assist the economy of Europe and prevent future war by integrating its members.

An influential paper by Martin, Mayer, and Thoenig (2008) argues that this conventional wisdom is only partially true and depends crucially on whether trade is bilateral or multilateral in nature. Bilateral trade increases the opportunity cost associated with the loss of trade between two countries and makes them therefore less likely to engage in war. In contrast, multilateral trade decreases bilateral dependence to any given country and thus the cost of bilateral conflict, making countries more likely to end up in war. The recent war in Ukraine and Russia's reliance on China to survive the trade disruption caused by sanctions imposed by the West fit this paradigm.

On the other hand, one might counter that in the case of Ukraine multilateralism had been severely compromised prior to the war and that it was in fact fractures in the world trading system that ultimately enabled the invasion of Ukraine. For instance, one might wonder if China would have been willing to provide a lifeline to Russia if its economic relationship with the West had not deteriorated so much in the past four years. The decline of multilateralism may not be evident in the trade data yet, though trade between the United States and China has declined following the trade war between these two countries. But trade data may take some time to adjust to reflect changes in attitudes and political climate that contribute to armed conflict.

The period leading up to World War II provides a case in point. De Bromhead and others (2019) study interwar trade and show that in the 1930s there was a dramatic shift away from multilateral trade toward trade within empires or informal spheres of influence. For instance, the British Empire accounted for around 30 percent of UK imports in 1929, but almost 42 percent by 1938. Several observers have argued that this changing nature of trade not only reflected, but potentially exacerbated, the international tensions of that period, an era now known as "pre-belligerency" (de Bromhead and others 2019). The changing nature of globalization in the past five years bears an eerie resemblance to that time.

IV. Conclusion

The future of globalization is highly uncertain at this point. But one thing is certain: there is no longer support for market-driven, unbridled globalization. Governments are now investing to reallocate international supply chains from a free market equilibrium. A challenge in this new era is that

such investments are motivated by objectives that are hard for researchers and the public to evaluate. When globalization was motivated primarily by the objective of economic efficiency, aggregate welfare benefits and distributional effects could be quantified through the lens of economic models. In contrast, there is not yet a quantitative benchmark for how much resilience is optimal. National security threats and their diminution can be difficult to verify without security clearance. Researchers can make progress by developing tools to evaluate the impacts of trade and industrial policy on these outcomes that are now guiding policymakers.

No matter what form globalization takes in the future, great care will have to be taken to address its potential effects on within-country inequality in advanced economies, the risks associated with high import concentration in critical product markets, and national security concerns. Whether industrial policy and trade restrictions aimed at containing China's technological development will accomplish this remains to be seen.

ACKNOWLEDGMENTS The authors thank Jan Gabriel Oledan for exceptional research assistance, and Jan Eberly, Pol Antràs, and Doug Irwin for helpful comments on an earlier draft.

References

Aiyar, Shekhar, Jiaqian Chen, Christian H. Ebeke, Roberto Garcia-Saltos, Tryggvi Gudmundsson, Anna Ilyina, and others. 2023. "Geoeconomic Fragmentation and the Future of Multilateralism." Staff Discussion Notes 2023/001. Washington: International Monetary Fund.

Antràs, Pol. 2020. "De-Globalisation? Global Value Chains in the Post-COVID-19 Age." Working Paper 28115. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28115.

Baldwin, Richard. 2019. *The Globotics Upheaval: Globalization, Robotics, and the Future of Work*. New York: Oxford University Press.

Barrot, Jean-Noël, and Julien Sauvagnat. 2016. "Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks." *Quarterly Journal of Economics* 131, no. 3: 1543–92.

Boehm, Christoph E., Aaron Flaaen, and Nitya Pandalai-Nayar. 2019. "Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tōhoku Earthquake." *Review of Economics and Statistics* 101, no. 1: 60–75.

Bonadio, Barthélémy, Zhen Huo, Andrei A. Levchenko, and Nitya Pandalai-Nayar. 2021. "Global Supply Chains in the Pandemic." *Journal of International Economics* 133:103534.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini. 2021. "Identity, Beliefs, and Political Conflict." *Quarterly Journal of Economics* 136, no. 4: 2371–411.

Breuninger, Kevin. 2021. "Pfizer CEO Opposes U.S. Call to Waive Covid Vaccine Patents, Cites Manufacturing and Safety Issues." CNBC, May 7. https://www.cnbc.com/2021/05/07/pfizer-ceo-biden-backed-covid-vaccine-patent-waiver-will-cause-problems.html.

Brunnermeier, Markus K. 2021. *The Resilient Society*. Colorado Springs, Colo.: Endeavor Literary Press.

Carvalho, Vasco M., Matthew Elliott, and John Spray. n.d. "Supply Chain Bottlenecks in a Pandemic." Work in Progress. https://covid.econ.cam.ac.uk/files/carvalho-files/BottlenecksPandemicNote.pdf.

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, and Alireza Tahbaz-Salehi. 2021. "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake." *Quarterly Journal of Economics* 136, no. 2: 1255–321.

Caselli, Francesco, Miklós Koren, Milan Lisicky, and Silvana Tenreyro. 2020. "Diversification through Trade." *Quarterly Journal of Economics* 135, no. 1: 449–502.

Casey, Christopher A. 2022. "Tariffs and the Infant Formula Shortage." Washington: Congressional Research Service. https://crsreports.congress.gov/product/pdf/IN/IN11932.

Cavallo, Alberto, Gita Gopinath, Brent Neiman, and Jenny Tang. 2021. "Tariff Pass-Through at the Border and at the Store: Evidence from US Trade Policy." *American Economic Review: Insights* 3, no. 1: 19–34.

Chari, Anusha, Peter Blair Henry, and Hector Reyes. 2021. "The Baker Hypothesis: Stabilization, Structural Reforms, and Economic Growth." *Journal of Economic Perspectives* 35, no. 3: 83–108.

Coppola, Antonio, Matteo Maggiori, Brent Neiman, and Jesse Schreger. 2021. "Redrawing the Map of Global Capital Flows: The Role of Cross-Border Financing and Tax Havens." *Quarterly Journal of Economics* 136, no. 3: 1499–556.

de Bromhead, Alan, Alan Fernihough, Markus Lampe, and Kevin Hjortshøj O'Rourke. 2019. "When Britain Turned Inward: The Impact of Interwar British Protection." *American Economic Review* 109, no. 2: 325–52.

De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. 2016. "Prices, Markups, and Trade Reform." *Econometrica* 84, no. 2: 445–510.

Dorn, David, and Peter Levell. 2021. "Trade and Inequality in Europe and the US." *IFS Deaton Review of Inequalities*, November 16. https://ifs.org.uk/inequality/trade-and-inequality-in-europe-and-the-us/.

Ellerbeck, Stefan. 2023. "What's the Difference between 'Friendshoring' and Other Global Trade Buzzwords?" World Economic Forum, February 17. https://www.weforum.org/agenda/2023/02/friendshoring-global-trade-buzzwords/.

Elliott, Matthew, Benjamin Golub, and Matthew V. Leduc. 2022. "Supply Network Formation and Fragility." *American Economic Review* 112, no. 8: 2701–47.

Evenett, Simon J. 2020. "Chinese Whispers: COVID-19, Global Supply Chains in Essential Goods, and Public Policy." *Journal of International Business Policy* 3:408–29.

Fajgelbaum, Pablo D., Pinelopi K. Goldberg, Patrick J. Kennedy, and Amit K. Khandelwal. 2020. "The Return to Protectionism." *Quarterly Journal of Economics* 135, no. 1: 1–55.

Fajgelbaum, Pablo D., Pinelopi K. Goldberg, Patrick J. Kennedy, Amit K. Khandelwal, and Daria Taglioni. 2022. "The US-China Trade War and Global Reallocations." Policy Research Working Paper 9894. Washington: World Bank. https://openknowledge.worldbank.org/entities/publication/68d193f0-815d-5cf9-8fcd-e6f41375ec09.

Farrell, Henry, and Abraham L. Newman. 2019. "Weaponized Interdependence: How Global Economic Networks Shape State Coercion." *International Security* 44, no. 1: 42–79.

Feldgoise, Jacob, and Remco Zwetsloot. 2020. "Estimating the Number of Chinese STEM Students in the United States." Issue Brief. Washington: Center for Security and Emerging Technology.

Goldberg, Pinelopi K. 2019. "The Future of Trade." *Finance and Development* 56, no. 2: 20–23.

Goldberg, Pinelopi K., with Greg Larson. 2023. *The Unequal Effects of Globalization*. Cambridge, Mass.: MIT Press.

Goldberg, Pinelopi K., and Nina Pavcnik. 2016. "The Effects of Trade Policy." In *Handbook of Commercial Policy, Volume 1A*, edited by Kyle Bagwell and Robert W. Staiger. Amsterdam: North-Holland.

Goldberg, Pinelopi K., and Tristan Reed. 2023. "Demand-Side Constraints in Development: The Role of Market Size, Trade, and (In)Equality." https://bpb-us-w2.wpmucdn.com/campuspress.yale.edu/dist/6/3741/files/2023/09/Demand_Side_Development.pdf.

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://nber.org/papers/w26918.

Irwin, Douglas A. 2019. "Does Trade Reform Promote Economic Growth? A Review of Recent Evidence." Working Paper 25927. Cambridge, Mass.: National Bureau of Economic Research. https://nber.org/papers/w25927.

Irwin, Douglas A. 2022. "The Trade Reform Wave of 1985–1995." NBER Working Paper 29973. Cambridge, Mass.: National Bureau of Economic Research. https://nber.org/papers/w29973.

Jia, Ruixue, Margaret E. Roberts, Ye Wang, and Eddie Yang. 2022. "The Impact of U.S.-China Tensions on U.S. Science." Working Paper 29941. Cambridge, Mass.: National Bureau of Economic Research. https://nber.org/papers/w29941.

Khanna, Gaurav, Nicolas Morales, and Nitya Pandalai-Nayar. 2022. "Supply Chain Resilience: Evidence from Indian Firms." Working Paper 30689. Cambridge, Mass.: National Bureau of Economic Research. https://nber.org/papers/w30689.

Kremer, Michael. 1993. "Population Growth and Technological Change: One Million B.C. to 1990." *Quarterly Journal of Economics* 108, no. 3: 681–716.

Li, Daitian, Tony W. Tong, and Yangao Xiao. 2021. "Is China Emerging as the Global Leader in AI?" *Harvard Business Review*, February 18. https://hbr.org/2021/02/is-china-emerging-as-the-global-leader-in-ai.

Martin, Philippe, Thierry Mayer, and Mathias Thoenig. 2008. "Make Trade Not War?" *Review of Economic Studies* 75, no. 3: 865–900.

Mattoo, Aaditya. 2018. "Services Globalization in an Age of Insecurity: Rethinking Trade Cooperation." Policy Research Working Paper 8579. Washington: World Bank.

McKinnon, John D., and Asa Fitch. 2022. "U.S. Restricts Semiconductor Exports in Bid to Slow China's Military Advance." *Wall Street Journal*, October 7. https://www.wsj.com/articles/u-s-restricts-semiconductor-exports-in-bid-to-slow-chinas-military-advance-11665155702.

Rauch, James E. 1999. "Networks versus Markets in International Trade." *Journal of International Economics* 48, no. 1: 7–35.

Rodríguez, Francisco, and Dani Rodrik. 2000. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." *NBER Macroeconomics Annual* 15:261–325.

Shapiro, Joseph S., and Reed Walker. 2018. "Why Is Pollution from US Manufacturing Declining? The Roles of Environmental Regulation, Productivity, and Trade." *American Economic Review* 108, no. 12: 3814–54.

Stumpner, Sebastian. 2022. "How Much Do China's Regional Lockdowns Affect International Trade?" Blog post, December 2, Eco Notepad, Banque de France. https://blocnotesdeleco.banque-france.fr/en/blog-entry/how-much-do-chinas-regional-lockdowns-affect-international-trade.

Sullivan, Jake. 2022. "Remarks at the Special Competitive Studies Project Global Emerging Technologies Summit," September 16. Washington: White House.

https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/09/16/remarks-by-national-security-advisor-jake-sullivan-at-the-special-competitive-studies-project-global-emerging-technologies-summit/.

Thun, Eric, Daria Taglioni, Timothy J. Sturgeon, and Mark Peter Dallas. 2022. "Massive Modularity: Understanding Industry Organization in the Digital Age—The Case of Mobile Phone Handsets." Policy Research Working Paper 10164. Washington: World Bank.

Ting-Fang, Cheng. 2022. "TSMC Founder Morris Chang Says Globalization 'Almost Dead.'" *Nikkei Asia*, December 7. https://asia.nikkei.com/Business/Tech/Semiconductors/TSMC-founder-Morris-Chang-says-globalization-almost-dead.

US Chamber of Commerce. 2021. *Understanding U.S.-China Decoupling: Macro Trends and Industry Impacts.* Washington: China Center, US Chamber of Commerce.

Varas, Antonio, Raj Varadarajan, Jimmy Goodrich, and Falan Yinug. 2021. *Strengthening the Global Semiconductor Supply Chain in an Uncertain Era*. Boston: Boston Consulting Group and Semiconductor Industry Association.

Wade, Adam, Anna Petherick, Beatriz Kira, Emily Cameron-Blake, Helen Tatlow, Jodie Elms, and others. 2023. "Variation in Government Responses to COVID-19." Working Paper 2023/032. Oxford: Blavatnik School of Government. https://www.bsg.ox.ac.uk/research/publications/variation-government-responses-covid-19.

White House. 2022. *National Security Strategy*. Washington. https://www.whitehouse.gov/wp-content/uploads/2022/10/Biden-Harris-Administrations-National-Security-Strategy-10.2022.pdf.

World Bank. 2020. *World Development Report 2020: Trading for Development in the Age of Global Value Chains.* Washington: World Bank.

World Bank. 2021. *Global Economic Prospects*. Washington: World Bank.

World Bank. 2023. *Global Economic Prospects*. Washington: World Bank.

Xie, Qingnan, and Richard B. Freeman. 2019. "Bigger Than You Thought: China's Contribution to Scientific Publications and Its Impact on the Global Economy." *China and World Economy* 27, no. 1: 1–27.

Yellen, Janet. 2022. "Resilient Trade." *Project Syndicate*, December 12. https://www.project-syndicate.org/magazine/biden-trade-agenda-emphasizes-resilience-by-janet-l-yellen-2022-12.

YouGov. 2017. "America's Friends and Enemies." February 2. https://today.yougov.com/topics/international/articles-reports/2017/02/02/americas-friends-and-enemies.

Comments and Discussion

COMMENT BY

POL ANTRÀS John Kenneth Galbraith once stated that "the only function of economic forecasting is to make astrology look respectable" (*Economist* 2016, par. 1). I am certainly more bullish than Galbraith was about the power of economics in delineating possible scenarios for the future of the world economy. But his irreverent assertion is a useful reminder that anything that is written on the future of globalization is necessarily speculative in nature. Although Goldberg and Reed's paper is no exception, their analysis constitutes an informed type of speculation, and a good illustration of the difference between a social science (economics) and a pseudo-science (astrology).

Goldberg and Reed make three main points in their article. First, they argue that the world economy does not appear to be deglobalizing; the hyperglobalization of the late 1980s, 1990s, and early 2010s has given way to a new era of "slowbalization," but there is little indication that globalization is in retreat. Second, and focusing on the recent COVID-19 pandemic, they show that global trade was remarkably resilient in the face of a drastic challenge to the functioning of the global economy. Finally, the authors identify a recent significant change in the policy environment and in public sentiment, particularly after Russia's invasion of Ukraine, a change that foretells a highly uncertain future for globalization, including scenarios reminiscent of the 1930s.

It will be no surprise to the authors, who have read and generously cite my own work on this topic (Antràs 2021), that I am largely in agreement with their main points. Regarding their first point about the lack of evidence for deglobalization, I am reminded of Robert Solow's famous

Brookings Papers on Economic Activity, Spring 2023: 397–423 © 2024 The Brookings Institution.

remark that "you can see the computer age everywhere but in the productivity statistics" (Solow 1987, par. 7). Similarly, one might say that you can see the deglobalization age everywhere but in international trade statistics. Unlike the case of Solow's productivity paradox, however, the disconnect between the widespread deglobalization rhetoric and actual data is not due to mismeasurement: as of today, the world is simply not deglobalizing. Furthermore, I also concur with the authors in their assessment that slowbalization has largely been caused by the growth of two giant economies, China and India, which have developed domestic capabilities that have allowed them to produce goods with less reliance on foreign intermediate inputs, thereby putting downward pressure on the ratio of world trade to world GDP.

Second, Goldberg and Reed's emphasis on the remarkable resiliency exhibited by world trade in the face of recent shocks, including the US-China trade war and the COVID-19 pandemic, is also uncontroversial. Some measures of globalization, such as the ratio of world trade to world GDP, plummeted in the spring of 2020, but the recovery from this collapse was fast and robust. Goldberg and Reed provide novel micro-level evidence supportive of the view that, if anything, trade integration increased the resilience of economies during the COVID-19 pandemic.

Third, I also agree that the rhetoric in policy circles appears to have taken a decidedly nationalistic tone in recent months. In fact, in my recent paper I concluded that "the main challenge for the future of globalisation is institutional and political in nature" and that the global pandemic that was unfolding at the time of my writing could "aggravate policy tensions across countries and further contribute to a new era of significant isolationism, much as the world witnessed in the 20th-century's Interwar Period" (Antràs 2021, 43). It may seem that I invoke these quotes to lobby for induction into the Astrologists' Hall of Fame, but I must come clean and admit that I did not foresee the geopolitical environment taking such a fast and such a worrisome turn for the worse. I was somewhat skeptical that President Biden would overturn many of President Trump's protectionist policies, but I was hopeful that he would have initiated talks with China, and I certainly did not expect the increasing diplomatic tensions between the United States and China that we have witnessed in March 2023. Similarly, I did not anticipate Russia's invasion of Ukraine in February 2022, though I will argue below that the implications of this conflict for the future of world trade are likely to be modest.

Despite my broad agreement with the authors' main arguments, I was asked to provide a critical discussion, so I must dutifully comply.

The remainder of this discussion will focus on outlining a few aspects in which my views depart a bit from those of the authors. In the process, I will also uncover some factors that the authors have ignored in their analysis but that I believe may be important in shaping the future of globalization.

RESILIENCY IS COSTLY I very much agree with Goldberg and Reed's emphasis on the fact that the relational nature of global value chains (GVCs) generates a form of stickiness that might explain why the geography of worldwide production has not yet changed much in the face of recent geopolitical shocks. Relatedly, the authors also acknowledge the role of sunk costs in generating this hysteresis, which is one of the leading themes of my 2021 paper on deglobalization.

One aspect that I would have perhaps stressed a bit more is that the sunk costs associated with the formation of global production chains are often very large in magnitude. To better grasp the relevance of these costs, consider the type of investments that a firm needs to carry out before being able to source parts, components, and services from a producer in a foreign country. First, the firm needs to gather information on a set of potential suitable suppliers in that country, or, in the case of greenfield investment, it needs to figure out a suitable location for a new plant. Next, the firm and its supplier need to invest in physical assets (a factory, specialized equipment capital, etc.) that are often customized to the needs of both parties. Finally, in an environment with imperfect contracting, the firm and its supplier will need to invest in relational capital, to ensure that the perceived contractual security of all agents in the transaction is sufficiently high. In practice, the latter implies that the initial transactions will be limited in size and will only grow slowly over time (Antràs and Foley 2015; Araujo, Mion, and Ornelas 2016).

An implication of these large overhead costs is that they naturally lead to the adoption of "lean and mean" global sourcing strategies. Although firms would prefer to rely on several suppliers to obtain a given component, in practice, multisourcing involves fixed costs that are too large for most firms to bear. Table 1 shows that only a very small share of US firms imports the same product (defined as a Harmonized System [HS] ten-digit code) from more than one source country.

In the face of recent events, it is tempting to argue that US firms were grossly under-diversified in their sourcing strategies and that, realizing their mistakes, they may seek more diversified strategies, such as a China Plus One approach. I am, however, skeptical about this view. Large global firms have sizeable procurement or supply chain management departments populated by highly trained managers. I think it is implausible to argue

Table 1. Number of Source Countries per Imported Product by US Firms

	Firm level		
	Mean	*Median*	*Max*
Mean	1.11	1.03	1.78
Median	1.00	1.00	1.00
95th percentile	1.61	1.00	4.00

Source: Reproduced from Antràs, Fort, and Tintelnot (2017), copyright American Economic Association.

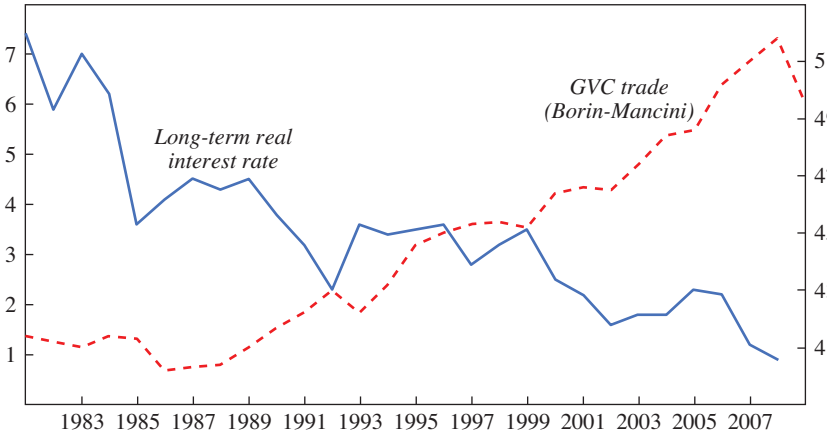
Note: Table reports statistics on the number of countries from which a firm imports the same HS10 product for the year 2007.

that these highly paid professionals systematically followed mistaken strategies, thereby inflicting millions of dollars in losses to their employers. Current supply chain disruptions will certainly lead firms to reassess their previous strategies, but my sense is that when firms crunch the numbers, they will realize just how expensive it is to bi-source or multisource, and thus most firms will abandon any drastic reorganization of their sourcing practices. As a result, we are not likely to see a significant increase in diversification in the near future.

THE ROLE OF INTEREST RATES A fact that trade economists often ignore is that the hyper-globalization period of the late 1980s, 1990s, and early 2000s came hand in hand with a substantial decline in real interest rates. Figure 1 depicts the share of GVC trade in gross exports based on a measure proposed by Borin and Mancini (2019), which corresponds to the share of world trade that involves more than one border crossing. By this metric, at the onset of the Great Recession, about 50 percent of world trade was related to GVC activity, with this share having climbed steadily in the previous twenty-eight years. The other series in figure 1 is borrowed from Farhi and Gourio (2018) and plots the ten-year US real interest rate (computed by subtracting inflation expectations from nominal Treasury yields) over the same period. The figure shows that US long-term real rates fell from about 7 percent in 1981 to less than 1 percent in 2009. A similar decline was observed in many other countries (Jordà and others 2019), and it not only applied to relatively risk-free rates but also to broader measures of the cost of capital faced by firms (Barkai 2020).

The causes of this secular decline in real interest rates are still being debated by macroeconomists, but candidate explanations include a slowdown in trend real output growth, demographic forces leading to an aging world population, a global saving glut, a shortage of safe assets, and increased wealth inequality.

Figure 1. Real Interest Rates and the Expansion of GVC (1981–2009)



Source: Author's compilation.

Note: The GVC trade line is created using a measure proposed by Borin and Mancini (2019); the long-term real interest rate line is adapted with permission from Farhi and Gourio (2018), copyright The Brookings Institution.

Independent of the causes of the observed fall in real interest rates, it seems plausible that this decline in the cost of capital was partly responsible for the remarkable growth in GVC activity. This is for a variety of reasons, which I have elucidated in recent work (Antràs 2023), but one salient aspect is that, as argued before, the formation of GVCs entailed significant setup costs, which were cheaper to finance in an era of low interest rates. Because many of these fixed costs were sunk in nature, a hypothetical reorganization of GVCs would again entail the need for significant new investments.

In an era of low or even negative real interest rates, this may not be a large impediment to reshoring or friendshoring, but the future path of real interest rates remains uncertain. As Goldberg and Reed point out, if an era of deglobalization is to play out, it is likely to lead to inflationary pressures. Even if those pressures are short-lived, they are likely to trigger tighter monetary policies around the globe, which may well increase the cost of borrowing for firms, and thereby make a reshoring or friendshoring less desirable.

In sum, even when diversification or reshoring may appear to be desirable strategies for firms, in the presence of large economies of scale in sourcing, it is far from obvious that they will be profit-maximizing policies when considering the sunk costs that would be associated with expanding

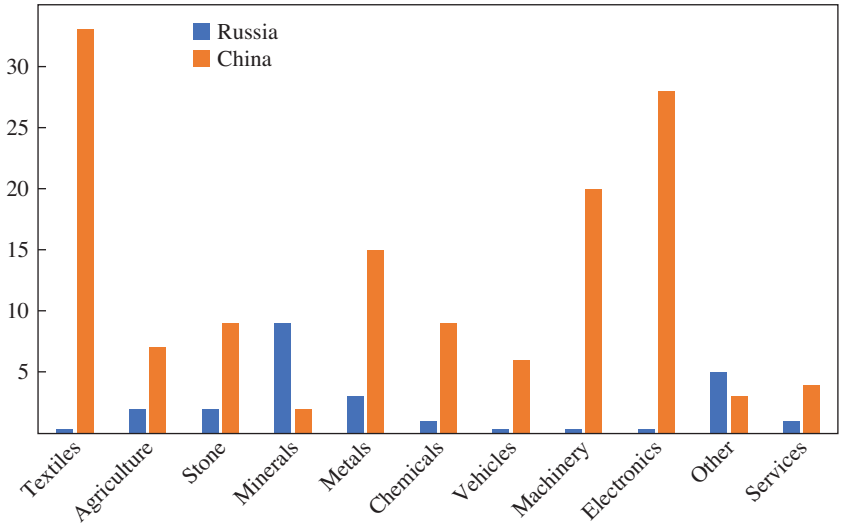
or reshuffling firms' global supply chains, particularly if firms contemplate these decisions in environments with higher interest rates than those that have prevailed in recent years.

RUSSIA VERSUS CHINA Russia's invasion of Ukraine in February 2022 constituted a flagrant violation of the territorial integrity of Ukraine, and the Russian government has justifiably been accused of war crimes and of crimes against humanity by several international organizations. Despite the dramatic aspects of this military conflict, I remain agnostic about its relevance for the future of globalization. Goldberg and Reed appear to confer much more relevance to this event, and they label the invasion a "catalyst for this new phase in the deglobalization movement." I do not deny that the Russia-Ukraine conflict has generated supply chain disruptions in the world economy (especially in Europe) and that it has also aggravated geopolitical tensions between the United States and China. Nevertheless, the implications of a decoupling of Russia from the global trade system would be relatively minor, and a counterfactual without the February 2022 invasion would have still witnessed rising geopolitical tensions.

To motivate my first point about the magnitude of the disruptions, it suffices to take a cursory look at the composition of the Russian economy as indicated by Russia's global share in the world exports of various (broadly defined) sectors. As is clear from figure 2, this share is negligible (lower than 0.5 percent) in all technologically intensive sectors in which GVC activity is dominant, such as vehicles, electronics, or machinery. Only in minerals and other primary products does Russia account for a significant share of world trade, but this share is lower than 10 percent for all two-digit industries. Even when focusing on narrower industry categories, such as exports of crude petroleum (HS code 2709) or natural gas (HS code 2711), the Russian market share is only 11.5 percent and 9.1 percent, respectively.[1] It is striking to contrast these figures with those recorded by China, as in figure 2. Clearly, a decoupling of China would have much dearer implications for the world economy than a Russian decoupling. This view is supported by the recent work of Bachmann and others (2022), who predicted that the economic effects of a potential cutoff of the German economy from Russian energy imports would not have been as catastrophic as some commentators in Germany were proclaiming.

1. See OEC, "Crude Petroleum," http://oec.world/en/profile/hs/crude-petroleum, and OEC, "Petroleum Gas," http://oec.world/en/profile/hs/petroleum-gas. It has been argued, however, that trade statistics do not provide an accurate portrait of trade in natural gas via pipelines. According to International Energy Agency statistics, Russia accounted for 18.7 percent of world exports of natural gas in 2020.

Figure 2. Russia's and China's Global Market Share in Exports by Sector



Source: The Growth Lab at Harvard University. The Atlas of Economic Complexity. http://www.atlas.cid.harvard.edu.

It may be argued that the shadow of the Cold War looms large and that Russia's invasion of Ukraine has rekindled geopolitical tensions that may drive the world toward increased isolationism. This appears to be what Goldberg and Reed suggest in their piece, particularly when they relate the invasion to increased concerns about national security, new demands of decoupling, the entry of the term "friendshoring" into the international trade vocabulary, and the use of export restrictions in the semiconductor sector. I am much less convinced about such a strong link existing. President Trump invoked national security concerns (via section 232 of the Trade Expansion Act of 1962) when he imposed tariffs on steel and aluminum in March 2018, close to four years before Russia's invasion. Similarly, concerns about the semiconductor industry, both in terms of a global chip shortage and the geopolitical consequences of the concentration of chip assembly in Taiwan, significantly predate the Russia-Ukraine crisis. Finally, the term "friendshoring" was popularized by US Secretary of the Treasury Janet Yellen in April 2022, but as Goldberg and Reed correctly point out, it had already been used in 2019 by US Commerce Secretary Gina Raimondo, when pushing Congress to approve a $52 billion package to bolster the US semiconductor sector (Shepardson 2021). For these

reasons, and as I mentioned above, it is not too clear to me that a counterfactual without the Russian invasion of Ukraine would have delivered a much rosier geopolitical environment today.

Regardless of the source of recent geopolitical tensions, I am increasingly concerned about the possibility of China decoupling from world markets. Until recently, I had held the view that the costs of decoupling (at least in present times) would have been too high for China to bear. A commitment to multilateral liberalization thus appeared to be the most pragmatic stance for China over the next few years. A key question, however, is with whom China would decouple. Most commentators and researchers take for granted that China would essentially go their own way, perhaps bringing some neighboring countries into their sphere of influence.[2] But the sheer size of its economy makes China the top trading partner for dozens of countries in the world, including many traditional allies of the United States, such as Australia, Japan, Korea, Chile, and Brazil.[3] If a decoupling scenario were to play out, these countries would face a nontrivial choice about which bloc to join, and the United States might end up being left more isolated than many commentators currently expect.

BEYOND GOODS MARKETS: DEGLOBALIZATION IN LABOR MARKETS Although most of Goldberg and Reed's paper is focused on the potential for a deglobalization of goods markets, they briefly comment on labor markets in section I. There, they point out that, despite a backlash against immigration in several parts of the globe, there is little evidence that migration flows have significantly slowed down or decreased in recent years. I share that opinion, and I provided similar evidence in my 2021 paper. Still, in recent years my views on this issue have become a bit more pessimistic. This is for at least four reasons.

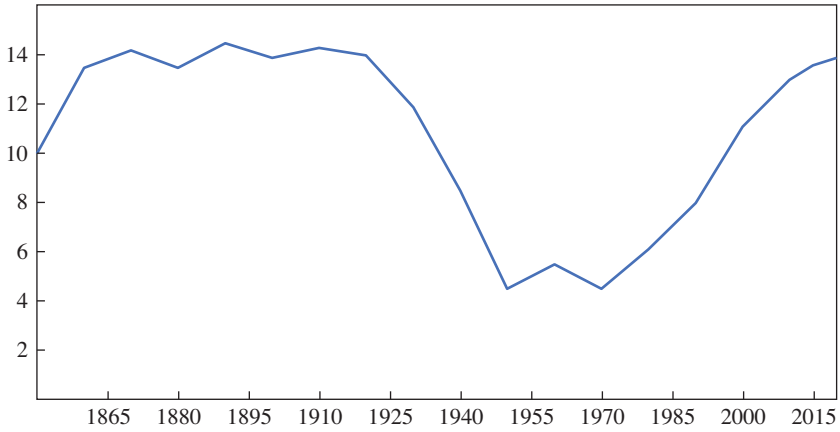
First, although the stock of migrants as a percentage of the world population has continued to rise in recent years, much of that increase is accounted for by a handful of countries that have welcomed refugees fleeing conflict. Many countries have been much less accommodating, and even in the United States, one can discern from figure 3 a slowdown (or even a halt) in the rate of growth of immigration.

Second, as figure 3 also indicates, periods of deglobalization in goods markets, such as the twentieth century's interwar period, tend to be periods

2. See, for example, Eppinger and others (2021).

3. See the illuminating figures at Visual Capitalist, "Countries Connected to Their Primary Trading Partner in 2020," https://www.visualcapitalist.com/wp-content/uploads/2022/02/2020-trading-partners.html.

Figure 3. Foreign-Born Population as a Percentage of the US Population (1850–2020)



Source: US Census Bureau.

with significant deglobalization in labor markets. Between 1920 and 1950, the share of foreign-born population in total US population fell from 14.8 percent to 4.7 percent. It is hard to predict whether the United States and other industrialized countries will implement tighter immigration policies in years to come, but we have already seen hints of increasingly strict policies. In the case of the United States, this has manifested itself in specific policies to curb illegal immigration of low-skilled individuals (such as President Trump's famous border wall with Mexico), but it has also translated into a severe shortage of highly educated foreign individuals, whom US firms in the tech industry would welcome with open arms (Shear and Jordan 2020).

Third, even if a scenario with significant deglobalization in labor markets is a low-probability event, it is important to emphasize that such a scenario would be exceedingly costly for the world economy. As a trade economist, I may lose sleep over the implications of a Chinese decoupling and the real income implications of the implied reorganization of GVCs. Yet, according to state-of-the-art quantitative trade models, the real income implications of such a decoupling would hardly amount to a few percentage points of world GDP.[4] Conversely, the real income implications of

4. For instance, Eppinger and others (2021) find that a repatriation of GVCs would reduce US real income by a mere 2.2 percent.

restricting migration flows are likely to be orders of magnitude larger, as forcefully argued by Clemens (2011). It may be argued that by focusing on static gains from trade, the current workhorse quantitative trade models underestimate the costs of decoupling in goods markets. For instance, a worse exploitation of economies of scale or diminished knowledge flows across countries may well have an impact on innovation and lead to magnified losses from deglobalization.[5] Nevertheless, the same exact argument can be made about quantitative evaluations of immigration restrictions. There is by now a significant body of work establishing a causal link between immigration and innovation, so the income costs of a deglobalization of labor markets are likely to remain orders of magnitude larger than those associated with a deglobalization in goods markets.[6]

The fourth and final reason for being particularly concerned about increasing restrictions on migration flows is the fact that the costs of those reduced flows are likely to be borne disproportionately by individuals who would like to emigrate but are not allowed to. These individuals would overwhelmingly have emigrated from less developed economies, and thus these restrictions will amount to leaving millions of individuals stuck with levels of income well below the world average. Under the plausible assumption that the marginal utility of income is decreasing in income, the welfare implications of migration restrictions are thus likely to be considerably larger than the associated real income implications.

NEEDED: A MULTIDISCIPLINARY APPROACH A famous adage, sometimes attributed to the physicist Niels Bohr and to others, including baseball legend Yogi Berra, states that "prediction is very difficult, especially if it is about the future."[7] Fortunately, economic science made enormous progress in the twentieth and early twenty-first centuries, so we are now much better equipped to make informed projections about the real income implications of differential possible policy scenarios.

This applies specifically to the field of international trade, which has become decidedly more empirical in recent years and has also embraced the quantitative revolution brought about by macroeconomics. Thanks to these

5. See, for instance, Sampson (2016).

6. See Kerr (2018) for a nontechnical overview of the link between innovation and immigration, and Bernstein and others (2022) for a more recent paper on this.

7. See, for example, Cranfield University, "Forecasting—Prediction Is Very Difficult, Especially If It's About the Future!," https://blogs.cranfield.ac.uk/cbp/forecasting-prediction-is-very-difficult-especially-if-its-about-the-future/.

developments, trade economists are now able to quickly produce projections for how world trade and world income will respond to counterfactuals with arbitrary changes in trade barriers. The work by Eppinger and others (2021) studying the implications of a Chinese "decoupling," by Bonadio and others (2021) and Çakmaklı and others (2021) on the response of world trade to the COVID-19 pandemic, or by Bachmann and others (2022) on the implications of Russia's invasion of Ukraine are recent examples of this exciting development.

As I have written elsewhere (Antràs and Chor 2021), I have some qualms about this quantitative revolution in the field of international trade. The most widely used models for quantitative analyses, such as the model by Caliendo and Parro (2015), constitute a blend of calibration and estimation, but for my taste, there is too high a ratio of calibration to estimation. More specifically, authors typically calibrate thousands of parameters (or combinations of parameters) to make their model fit the data exactly. This, I believe, is problematic because these models impose strong assumptions on functional forms; calibrating thousands of parameters to fit the data exactly amounts to a form of overfitting that is likely to lead to poor out-of-sample performance.

In my view, the international trade field should transition away from high-scale calibration exercises and fully embrace more standard structural estimation approaches, which are pervasive in many other areas in economics (such as industrial organization), and which are likely to produce more reliable counterfactuals.

As a matter of fact, I will conclude with a call for even more interdisciplinary studies of the world economy. Trade economists should look not just outside of trade or macroeconomics and borrow from other fields in economics. But they should also look beyond economics. When asked to speculate about the future of the world economy, trade economists are equipped to trace the real income implications of different geopolitical scenarios, but we have little sense of which geopolitical scenarios are more likely to play out. To the question, Will the world economy deglobalize? we can at best play around with different scenarios in which trade barriers vis-à-vis certain trading partners go up by different possible amounts. I firmly believe that these exercises are useful and can have an impact on policy, but they are not necessarily useful in producing a meaningful probability distribution over possible future scenarios.

Given the prominence of geopolitical forces shaping current events, it seems particularly important for trade economists to gain a better

understanding of the interplay between international relations and international trade. In our quantitative trade models, governments set taxes and tariffs, and they may even engage in trade wars and trade talks, but the modeling and objectives of these governments are quite distinct from how international relations scholars view governments as players trying to flex their power under the Hobbesian anarchy of the international arena. The geopolitical environment shapes economic policy and results in a particular configuration of international trade flows. But the structure of international trade flows also affects the payoffs of governments in the underlying geopolitical game, even if these governments care about more than the mere economic well-being of their constituents. Only by jointly modeling these forces will trade economists be able to produce well-defined projections for how the world economy is likely to evolve over the coming years.

REFERENCES FOR THE ANTRÀS COMMENT

Antràs, Pol. 2021. "De-globalisation? Global Value Chains in the Post-COVID-19 Age." Working Paper 28115. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28115.

Antràs, Pol. 2023. "An 'Austrian' Model of Global Value Chains." Working Paper 30901. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30901.

Antràs, Pol, and Davin Chor. 2021. "Global Value Chains." In *Handbook of International Economics: International Trade, Volume 5*, edited by Elhanan Helpman, Gita Gopinath, and Kenneth Rogoff. Amsterdam: North-Holland.

Antràs, Pol, and C. Fritz Foley. 2015. "Poultry in Motion: A Study of International Trade Finance Practices." *Journal of Political Economy* 123, no. 4: 853–901.

Antràs, Pol, Teresa C. Fort, and Felix Tintelnot. 2017. "The Margins of Global Sourcing: Theory and Evidence from US Firms." *American Economic Review* 107, no. 9: 2514–64.

Araujo, Luis, Giordano Mion, and Emanuel Ornelas. 2016. "Institutions and Export Dynamics." *Journal of International Economics* 98:2–20.

Bachmann, Rüdiger, David Baqaee, Christian Bayer, Moritz Kuhn, Andreas Löschel, Benjamin Moll, Andreas Peichl, Karen Pittel, and Moritz Schularick. 2022. "What If? The Economic Effects for Germany of a Stop of Energy Imports from Russia." EconPol Policy Report 36. Munich: CESifo.

Barkai, Simcha. 2020. "Declining Labor and Capital Shares." *Journal of Finance* 75, no. 5: 2421–63.

Bernstein, Shai, Rebecca Diamond, Abhisit Jiranaphawiboon, Timothy McQuade, and Beatriz Pousada. 2022. "The Contribution of High-Skilled Immigrants to Innovation in the United States." Working Paper 30797. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30797.

Bonadio, Barthélémy, Zhen Huo, Andrei A. Levchenko, and Nitya Pandalai-Nayar. 2021. "Global Supply Chains in the Pandemic." *Journal of International Economics* 133:103534.

Borin, Alessandro, and Michele Mancini. 2019. "Measuring What Matters in Global Value Chains and Value-Added Trade." Policy Research Working Paper 8804. Washington: World Bank Group. https://documents1.worldbank.org/curated/en/639481554384583291/pdf/Measuring-What-Matters-in-Global-Value-Chains-and-Value-Added-Trade.pdf.

Çakmaklı, Cem, Selva Demiralp, Şebnem Kalemli-Özcan, Sevcan Yeşiltaş, and Muhammed A. Yıldırım. 2021. "The Economic Case for Global Vaccinations: An Epidemiological Model with International Production Networks." Working Paper 28395. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28395.

Caliendo, Lorenzo, and Fernando Parro. 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *Review of Economic Studies* 82, no. 1: 1–44.

Clemens, Michael A. 2011. "Economics and Emigration: Trillion-Dollar Bills on the Sidewalk?" *Journal of Economic Perspectives* 25, no. 3: 83–106.

Economist. 2016. "A Mean Feat." January 9. https://www.economist.com/finance-and-economics/2016/01/09/a-mean-feat.

Eppinger, Peter, Gabriel J. Felbermayr, Oliver Krebs, and Bohdan Kukharskyy. 2021. "Decoupling Global Value Chains." Working Paper 9079. Munich: CESifo. https://www.cesifo.org/en/publications/2021/working-paper/decoupling-global-value-chains.

Farhi, Emmanuel, and François Gourio. 2018. "Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia." *Brookings Papers on Economic Activity*, Fall, 147–223.

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor. 2019. "The Rate of Return on Everything, 1870–2015." *Quarterly Journal of Economics* 134, no. 3: 1225–98.

Kerr, William R. 2018. *The Gift of Global Talent: How Migration Shapes Business, Economy and Society*. Stanford, Calif.: Stanford University Press.

Sampson, Thomas. 2016. "Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth." *Quarterly Journal of Economics* 131, no. 1: 315–80.

Shear, Michael D., and Miriam Jordan. 2020. "Trump Suspends Visas Allowing Hundreds of Thousands of Foreigners to Work in the U.S." *New York Times*, June 22. https://www.nytimes.com/2020/06/22/us/politics/trump-h1b-work-visas.html.

Shepardson, David. 2021. "U.S. Commerce Chief Urges Congress to Act on Chip Funding by August." Reuters, June 28. https://www.reuters.com/technology/us-commerce-chief-urges-congress-act-chip-funding-by-august-2021-06-28/.

Solow, Robert M. 1987. "We'd Better Watch Out." Review of *Manufacturing Matters: The Myth of the Post-Industrial Economy*, by Stephen S. Cohen and John Zysman. *New York Times Book Review*, July 12.

COMMENT BY

DOUGLAS A. IRWIN Twenty-four years ago, at the Brookings Trade Forum, I presented a paper entitled "Is Globalization Today Really Different from Globalization a Hundred Years Ago?" (Bordo, Eichengreen, and Irwin 1999). That paper was written at a time of rapidly increasing global economic integration. The opening of China, the transition from communism to capitalism in Eastern Europe and the former Soviet Union after the fall of the Berlin Wall, and the reduction of trade and investment barriers by many developing countries during the 1980s and 1990s all contributed to a remarkable expansion of world commerce.

The purpose of the paper was to evaluate whether world economic integration at the time had exceeded that experienced in the first era of globalization in the late nineteenth and early twentieth centuries. That earlier globalization saw steadily increasing trade flows, large international capital movements under the gold standard, and the mass migration of Europeans to North and South America and elsewhere. By looking at basic data on trade volumes, capital flows, and labor migration during the two periods, we concluded that the mid-1990s era of globalization had gone beyond that seen prior to World War I. At the time, that was a closer call than you might have thought because global integration was pretty extensive by 1914. In the decade after our paper was written, of course, globalization intensified into what some called the hyper-globalization of the 2000s. That left no doubt that early twenty-first-century integration had surpassed that of the early twentieth century.

Thinking back to 1999, it would have been difficult to imagine that just a quarter of a century later Brookings would have to have a discussion about deglobalization. Or would it? Around the time our paper was written, several economists, such as Jeffrey Williamson (1998), had forebodings about the future and reminded us that what could go up could also go down. After all, the first era of globalization crashed and burned with the outbreak of World War I: the gold standard was abandoned, trade barriers were raised, and immigration restrictions were imposed. The interwar period of the 1920s and 1930s saw deglobalization on a large scale—less trade, reduced migration, and diminished capital flows. So reversals of globalization are certainly possible, and their effects can be long-lasting. It took several decades after World War II for the world economy to recover from that costly implosion.

The current era of globalization has not run into a world war, but perhaps it has hit a brick ceiling. As Goldberg and Reed note in their fine paper, globalization seems to have peaked around the time of the global

financial crisis of 2008–2009 and has encountered many difficulties since. The rise of geopolitical tensions, leading to the US-China trade war, questions about the resilience of supply chains during the COVID-19 pandemic, increasing national security concerns about excessive dependence on semiconductors from Asia, disruptions caused by Brexit and Russia's invasion of Ukraine, talk of friendshoring in trade and investment, the use of government regulation (and tax incentives) to promote domestic production of electric vehicles, and the introduction of a carbon border adjustment mechanism as part of the European Union's environmental policy—all suggest we have entered a new era.

Goldberg and Reed have done an excellent job of providing an early assessment of this new period. They have a difficult assignment because events are changing rapidly and government actions are a moving target. Many policy measures have been proposed and some are about to be implemented, so their long-term impact on trade and investment flows has yet to be seen. Still, the paper has many insightful and wise observations. Their basic point, which they argue effectively, is that the world is not deglobalizing. Instead, government policies—in some sectors and with some countries—are acting to brake if not reverse globalization on grounds of resilience, national security, and the environment.

Globalization is often described as the increased interaction between countries in terms of trade in goods and services, capital flows, labor migration, and technology and data. The paper does not take a comprehensive view of globalization in each of these areas but rightly focuses on trade in goods.[1] That is where policy action, particularly in the United States, appears to be focused.[2] In doing so, they push back against some of the common narratives that appear to be driving policy. For example, they conclude that "despite the prominence of resilience concerns in the public debate in the past three years the evidence to date provides no support either for the view that global supply chains were not resilient during the pandemic or that the world economy would have been more resilient if there had been less dependence on foreign inputs and trade." They maintain that the policy backlash is not being driven by producer or worker interests but by governments: "deglobalization is not market-driven; rather it is the result of government-led actions and policies that had little support, at least initially, from the private sector."

1. This is something that Chandy and Seidel (2016) did a few years ago.

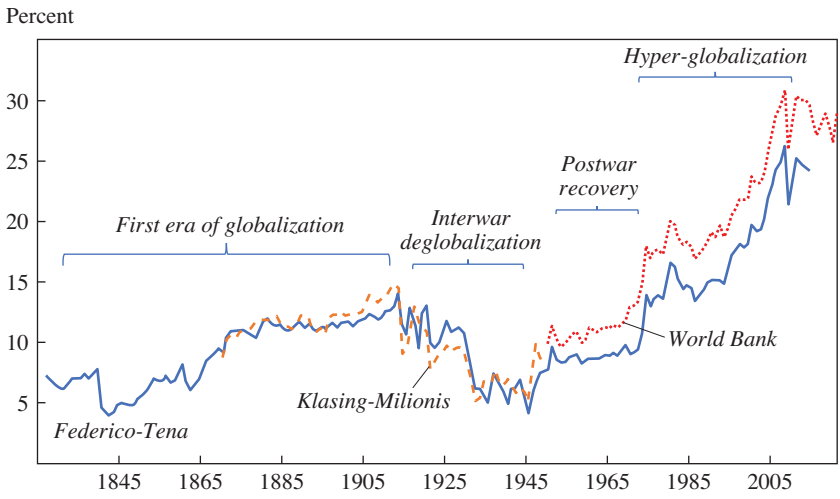
2. Of course, the Trump administration tried to reduce immigration, while seeking to encourage foreign capital inflows and discourage capital outflows.

They also note that "national security is the most powerful argument against unconstrained, market-driven globalization to date. It is also the most problematic as it cannot be verified directly by researchers, market analysts, or journalists—one has to place faith in the government's intelligence sources." This important observation has implications for all trade policy economists. Efficiency or even distributional effects of policies are no longer the principal criteria used to evaluate the desirability of many policy actions. Instead, national security considerations now dominate the discussion. (Foreign policy considerations, of course, have always been paramount in actual decision-making, although in the past they were often aligned with efficiency goals.) Ever since Adam Smith famously stated that "defence, however, is of much more importance than opulence" (1976, 464–65), economists usually step back, concede that security trumps efficiency, and accept the prima facie case for national security. Such interventions in trade are usually taken as given and put in the bucket of noneconomic objectives that governments have, as constraints to be minimized in the quest for economic efficiency (Bhagwati 1967; Hoekman, Mavroidis, and Nelson 2023). But the stakes are so high and the national security claims have become so broad and all-encompassing that economists have to get involved in this debate. Unfortunately, national security considerations are hard for economists to evaluate, often requiring evaluation of tail risks and other factors that defy easy measurement or quantification. They add a big element of uncertainty and difficult judgment calls into standard cost-benefit calculations.

Since I largely agree with the thrust of the paper, I thought I would put recent globalization trends in historical perspective rather than comment on the specific contributions of this paper. Figure 1 of the paper shows the ratio of world exports to GDP since 1989, where there is a flatlining of the share since about 2008. The flattening seems to indicate a break from the previous twenty years, but is it unusual? Is it a historical anomaly, or have we seen such pauses in the past?

To answer this question, we need a longer time series. Figure 1 here presents the ratio of world trade to GDP going back to 1827. With this figure we can identify five stages of modern globalization. The first stage, from 1827 to 1914, is the first era of globalization and the later part of this period served as a benchmark for Bordo, Eichengreen, and Irwin (1999). This was a period of rising trade due to lower transportation costs, with the introduction of the steamship and refrigeration, and some European tariff reductions following the Cobden-Chevalier trade agreement between Britain and France in 1860. This was a period of growing capital flows under the gold standard and mass migration from Europe to the rest of the world.

Figure 1. World Exports as a Share of World GDP, 1827–2021



Sources: Data from Klasing and Milionis (2014); Federico and Tena-Junguito (2019); and World Bank.
Note: Data available at https://www.uc3m.es/ss/Satellite/UC3MInstitucional/es/TextoMixta/1371246237481/Federico-Tena_World_Trade_Historical_Database and https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS.

Toward the end of this era of globalization there were rising geopolitical tensions between the old superpower of Britain and the rising industrial power of Germany. Rising trade frictions also led to a modest protectionist backlash in the late nineteenth century, particularly in agricultural goods but also to some extent for manufactured goods (Williamson 1998). These factors may account for the stabilization of the export-to-GDP ratio after 1880, when this measure of globalization seems to stall.

The second stage, from 1914 to 1945, encompasses two destructive world wars and the intense protectionism of the Great Depression of the 1930s. This period of deglobalization is characterized by higher trade barriers, more capital controls, and immigration restrictions. A third stage covers the postwar period from 1945 to the mid-1980s when trade recovers but without participation of the second world (the communist countries of the Soviet Union, China, and Eastern Europe) or the third world (the Global South, which was committed to import substitution policies). When those countries started to rejoin the global economy in the 1980s and 1990s, economic integration really took off, marking a fourth stage. The question is whether we have entered a new, fifth stage of deglobalization since 2008.

How should we understand these globalization cycles? Goldberg and Reed suggest that integration is driven by changes in technology and in policy. Just as Goldin and Katz (2010) suggest that there is a race between technology and education in determining the wage premium for higher-educated workers, there is a similar interplay between technology and policy in the case of trade integration. Falling trade costs due to technological developments can act as an accelerator for global integration in bringing markets together.[3] By contrast, policy can work either to bring markets together, such as deepening integration through trade agreements, or push them further apart by raising trade barriers across markets. (To use a car analogy, policy can act not just as a brake on integration but can also throw the whole vehicle into reverse, as seen in the interwar period.)

This analogy helps us interpret the history of global integration but also think about the past few decades. The hyper-globalization period of the 1990s and 2000s was a period in which technology and policy were working in the same direction, toward the integration of markets. New innovations such as the shipping container and air transport were fully exploited in moving goods across markets. New trade agreements in the early 1990s, such as the North American Free Trade Agreement (NAFTA) and the Uruguay Round of trade negotiations that created the World Trade Organization (WTO), brought down trade barriers and hence trade costs (World Bank 2020). Equally if not more important were unilateral reforms in many emerging markets, such as China and India (Irwin 2022).

Today, technologies that reduce trade costs (at least for moving merchandise goods) may have hit diminishing returns, slowing global integration except in places such as Africa, where the scope for further improvements is great. Meanwhile, trade liberalization has stalled both at the multilateral level (no major agreements at the WTO) and at the unilateral level. Even worse, policy has moved in a reverse direction and is pulling markets apart, as new trade restrictions creep into the system. As Goldberg and Reed note, this policy shift from previous decades is being driven by governments of their own accord, which are not acting at the behest of producers and workers who are adversely affected by foreign competition. The US foreign policy establishment, which had long supported trade integration as enhancing national security, now views it differently. The Trump and Biden administrations have both rejected new trade agreements on grounds of national security and worker welfare, marking a new path for US trade policy.

3. Of course, some technologies can substitute domestic production for imports or external sourcing, such as 3D printing.

It is difficult to predict whether global integration simply stabilizes at current levels, as it did previously, in the 1880s and 1960s, or actually declines, as it did in the interwar period. At this point it is hard to imagine the configuration of technology and policy that will push global integration to much higher levels. All this depends on future policy developments.

Goldberg and Reed are optimistic that we are not on the cusp of a deglobalization era. Richard Baldwin (2022) gives us further reasons to believe that the recent decline in the trade share is not a cause for worry. He finds that about two-thirds of the slide is due to the declining value of trade in mining and fuels. Probing further, he finds that this slide is not due to falling volumes but to falling prices. As metals and fuel prices came off a historic high from 2011, marking the end of a commodity super cycle, the value of world trade in these categories has fallen to more normal levels. Thus, the declining trade share is partly a benign development not driven by policy decisions.

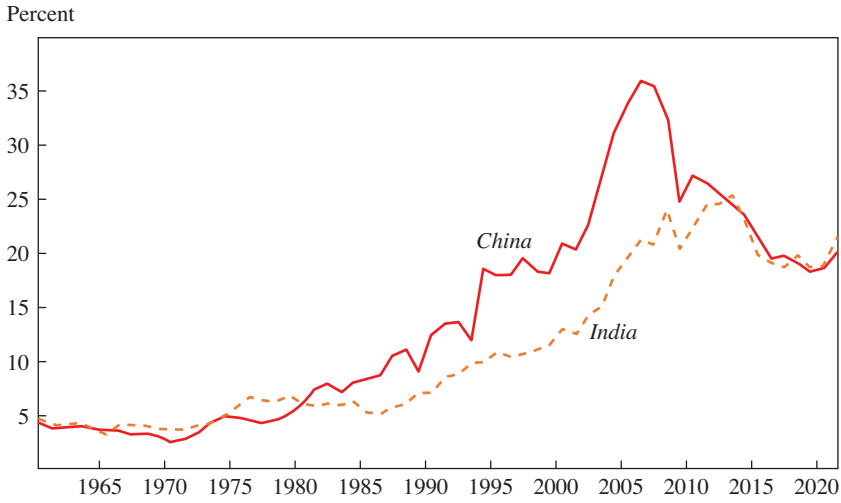
Baldwin (2022) and others also point out that trade in services and data is still growing at a healthy rate and shows no signs of faltering. While these flows are still small in relation to merchandise trade, these developments suggest that globalization is evolving and changing, not declining.

At the same time, it is worth noting that part of the decline in the export-to-GDP ratio can be observed in two large countries, China and India. As figure 2 shows, both countries are coming off large increases in the export-to-GDP ratios in the 2000s. Whether the recent decline is due to some natural rebalancing or to active policy measures to limit trade is unclear and deserves further study. But a case can be made that these two countries have been turning inward under President Xi Jinping and Prime Minister Narendra Modi.

Should we be concerned about the slight reversal of this measure of trade integration? Of course, any answer depends on the reasons for the decline, but if driven by government policies, then there are welfare consequences to any potential deglobalization scenario. If global economic integration led to tangible gains, then disintegration will lead to economic losses.[4] The International Monetary Fund has collected some preliminary evidence on the welfare cost of geopolitical fragmentation of trade and suggests that it could be around 1 percent of GDP in a situation limited to trade fragmentation. If there was to be full technological decoupling and reduced knowledge diffusion combined with sectoral misallocation, the costs are much higher, in the 8–10 percent range (Aiyer and others 2023). While the

4. For evidence on the case of Brexit, see Dhingra and Sampson (2022).

Figure 2. Exports as a Share of GDP, China and India, 1960–2021



Source: World Bank.
Note: Data available at https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS?locations=CN.

precise estimates vary greatly depending on the model assumptions and scenario envisioned, a fragmentation of the world economy could come at a substantial price.

Goldberg and Reed's paper will stimulate further work on these matters. The big question is how much government policies will actively try to further reverse the high level of integration achieved over the past few decades.

REFERENCES FOR THE IRWIN COMMENT

Aiyar, Shekhar, Jiaqian Chen, Christian H. Ebeke, Roberto Garcia-Saltos, Tryggvi Gudmundsson, Anna Ilyina, and others. 2023. "Geoeconomic Fragmentation and the Future of Multilateralism." Staff Discussion Notes 2023/001. Washington: International Monetary Fund.

Baldwin, Richard. 2022. "The Peak Globalisation Myth." Blog post, November 5, BaldwinRE Substack. https://rbaldwin.substack.com/p/the-peak-globalisation-myth.

Bhagwati, Jagdish. 1967. "Non-economic Objectives and the Efficiency Properties of Trade." *Journal of Political Economy* 75, no. 5: 738–42.

Bordo, Michael D., Barry Eichengreen, and Douglas A. Irwin. 1999. "Is Globalization Today Really Different from Globalization a Hundred Years Ago?" *Brookings Trade Forum*, 1–72. http://www.jstor.org/stable/25063137.

Chandy, Laurence, and Brina Seidel. 2016. *Is Globalization's Second Wave about to Break?* Global Views, no. 4. Washington: Brookings Institution.

Dhingra, Swati, and Thomas Sampson. 2022. "Expecting Brexit." *Annual Review of Economics* 14:495–519.

Federico, Giovanni, and Antonio Tena-Junguito. 2019. "World Trade, 1800–1938: A New Synthesis." *Revista de Historia Económica—Journal of Iberian and Latin America Economic History* 37, no. 1: 9–41. https://doi.org/10.1017/S0212610918000216.

Goldin, Claudia, and Lawrence F. Katz. 2010. *The Race between Education and Technology*. Cambridge, Mass.: Harvard University Press.

Hoekman, Bernard M., Petros C. Mavroidis, and Douglas R. Nelson. 2023. "Geopolitical Competition, Globalisation and WTO Reform." *World Economy* 46, no. 5:1163–88.

Irwin, Douglas A. 2022. "The Trade Reform Wave of 1985–1995." *American Economic Association Papers and Proceedings* 112:244–51.

Klasing, Mariko J., and Petros Milionis. 2014. "Quantifying the Evolution of World Trade, 1870–1949." *Journal of International Economics* 92, no. 1: 185–97.

Smith, Adam. 1976. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford: Clarendon Press.

Williamson, Jeffrey G. 1998. "Globalization, Labor Markets and Policy Backlash in the Past." *Journal of Economic Perspectives* 12, no. 4: 51–72.

World Bank. 2020. *World Development Report 2020: Trading for Development in the Age of Global Value Chains.* Washington: World Bank.

GENERAL DISCUSSION Maurice Obstfeld began with a comment on the geopolitical implications of globalization. Obstfeld offered China's Belt and Road Initiative and its Regional Comprehensive Economic Partnership as examples of a country using trade to build regional relationships. He noted that the United States had started down an analogous route, hoping that the Trans-Pacific Partnership (TPP), alongside its trade benefits, would promote US geopolitical influence in Asia and over the rules of the broader world trading system.[1] Obstfeld suggested that US advocacy of friendshoring, or the act of primarily trading with and investing in ally countries,

1. Mireya Solis, "The Geopolitical Importance of the Trans-Pacific Partnership: At Stake, a Liberal Economic Order," Brookings Commentary, March 13, 2015, https://www.brookings.edu/articles/the-geopolitical-importance-of-the-trans-pacific-partnership-at-stake-a-liberal-economic-order/; Daniel Twining, Hans Kundnani, and Peter Sparding, *Trans-Pacific Partnership: Geopolitical Implications for EU-US Relations* (Strasbourg: European Parliament, 2016), https://www.europarl.europa.eu/RegData/etudes/STUD/2016/535008/EXPO_STU(2016)535008_EN.pdf; and Cary Huang, "It's the Geopolitics, Stupid: US-Led TPP Trade Pact Less about Boosting Economies Than about Containing China's Rise," *South China Morning Post*, November 6, 2015, https://www.scmp.com/news/china/policies-politics/article/1876024/its-geopolitics-stupid-us-led-tpp-trade-pact-less-about?campaign=1876024&module=perpetual_scroll_0&pgtype=article.

would be perceived quite differently now had the United States followed through on TPP in 2016–2017, rather than turning toward protection. In particular, he argued, a less protectionist United States embedded in TPP might have had more leverage to convince middle-income countries to support Western sanctions on Russia after its invasion of Ukraine. Obstfeld argued that in current circumstances, friendshoring reinforces negative sentiment and tense relationships between countries. He suggested that it perpetuates a cycle of fragmentation in the world economy.

Obstfeld also stated that current metrics of financial globalization are insufficient. He argued that the marginal financial trade between richer countries likely has minimal beneficial welfare effects and that there are instances, such as tax arbitrage or during the housing bubbles of the 2000s, where further financial globalization actually has negative social impacts. In contrast, Obstfeld said, the marginal capital flow to emerging markets has more potential to be socially productive, but these flows tend to be insufficient. Thus, typical quantitative measures of financial globalization trends may not be a sound basis for judgments on the benefits coming from cross-border asset trade.

Tristan Reed responded to Obstfeld and referred to figure A1 in the online appendix, which uses a time series by Coppola and others.[2] The figure accounts for foreign investment through tax havens and still shows an increase in financial trade.

Viral Acharya asked, in reference to the previous day's discussion on emerging markets, if changing sentiment and rhetoric on globalization is related to income inequality within countries. He continued by asking if the authors' empirical findings on globalization could be linked with macroeconomic outcomes such as widening inequality. Acharya also pointed out the discrepancy between research in manufacturing and goods versus services. He wondered if researchers' continual concern with manufacturing is a result of the jobs it creates. He added that the manufacturing and goods sector typically creates jobs on the lower end of the income distribution, whereas jobs in the professional and financial services are traditionally higher.[3] He suggested these differences may contribute to the changes in sentiment toward free trade.

2. Antonio Coppola, Matteo Maggiori, Brent Neiman, and Jesse Schreger, "Redrawing the Map of Global Capital Flows: The Role of Cross-Border Financing and Tax Havens," *Quarterly Journal of Economics* 136, no. 3 (2021): 1499–556.

3. US Bureau of Labor Statistics, "Table B-3. Average Hourly and Weekly Earnings of All Employees on Private Nonfarm Payrolls by Industry Sector, Seasonally Adjusted," https://www.bls.gov/news.release/empsit.t19.htm.

John Haltiwanger elaborated Acharya's point on the distinction between goods and services research, commenting on the discrepancy in data quality between the two sectors. He suggested research may be understating the growth in trade in services and that we could be witnessing a restructuring, not necessarily deglobalization. He added that it would be interesting to think about the changing political climate in relation to this restructuring. He continued by saying recent conversations are centered around the emergence and growth potential of artificial intelligence (AI). Haltiwanger said trade in services may grow if these predictions are correct.

Pinelopi Goldberg agreed that trade in services will become more important. However, she added, today it occurs on a much smaller scale and in a more restrictive way than do trade in goods.[4] She pointed to Douglas Irwin's discussion, in which he linked globalization and hyperglobalization to technology *and* policy. Although there are technological advancements in AI, Goldberg argued that policy around AI is moving in the opposite direction. She claimed that the current policy landscape does not support an increase in trade in services, but acknowledged that could change at any time.

Anne Case expanded on Acharya's points by mentioning the NAFTA negotiations. She stated that economists were aware that NAFTA would have a negative impact on the local labor market but presented it as an opportunity to retrain and reskill workers.[5] Case argued that because retraining never took place, this resulted in job losses that generated resentment toward international production and labor.[6] She added that, while important, COVID-19 and Russia's impact on supply chains only alleviated pressure that had been building since NAFTA.

Joseph Gagnon raised an additional point on trade imbalances. He cited work from the Centre for Economic Policy Research that found a strong, predictive relationship between trade imbalances and protectionist policies.[7] Gagnon further mentioned Goldberg and Reed's finding that the

4. OECD, "Trade in Goods and Services," https://data.oecd.org/trade/trade-in-goods-and-services.htm.

5. Shushanik Hakobyan and John McLaren, "Looking for Local Labor Market Effects of NAFTA," *Review of Economics and Statistics* 98, no. 4 (2016): 728–41, https://www.jstor.org/stable/24917047.

6. Public Citizen's Global Trade Watch, "NAFTA's Legacy: Lost Jobs, Lower Wages, Increased Inequality," factsheet, October 2019, https://www.citizen.org/wp-content/uploads/NAFTA-Factsheet_Deficit-Jobs-Wages_Oct-2019.pdf.

7. Etienne Fize, Philippe Martin, and Samuel Delpeuch, "Trade Imbalances and the Rise of Protectionism," VoxEU, February 12, 2021, https://cepr.org/voxeu/columns/trade-imbalances-and-rise-protectionism.

United States and the United Kingdom have the largest trade deficits.[8] He reflected on Donald Trump's own trade policies and his focus on closing trade deficits, despite them growing during his administration.[9] Gagnon suggested using the relationship between deficits and protectionist policies as a framework for discussing deglobalization. He used foreign exchange intervention and capital inflow taxes as examples of how policy could reduce deficits and minimize tensions between nations.

Eswar Prasad commented on the increased risk corporations with large-scale, multinational supply chains face. He reflected on the difficulties and high financing costs corporations confront when augmenting their supply chains in response to risk. Prasad suggested looking at other evidence to track deglobalization, such as foreign direct investment (FDI) flows. He referenced research from the International Monetary Fund (IMF) on geo-economic fragmentation.[10] The paper found a positive relationship between FDI flows and geopolitically aligned countries. However, Prasad used local shocks as an example of how realignment may not increase resilience. Although a corporation adapting its supply chain may be inefficient, Prasad noted that there may be an added insurance value for corporations knowing its supply chains are in a geopolitically aligned country.

Prasad continued by discussing the impact protectionist policies will have on low-income countries and global inequality. He used the Inflation Reduction Act and the CHIPS and Science Act as two examples of the United States prioritizing domestic production. Prasad referred to the Made in China 2025 and the Made in India initiatives as examples of similar international policies reflecting protectionist sentiment. He argued that these policies will exacerbate global inequality because they leave out economies that never benefited from globalization.

Reed agreed with Prasad's concerns about global inequality and cited recent evidence that trade is being funneled between high-income countries

8. Bureau of Economic Analysis, "U.S. International Trade in Goods and Services, February 2023," news release, https://www.bea.gov/news/2023/us-international-trade-goods-and-services-february-2023; Office for National Statistics, "UK Trade: January 2023," https://www.ons.gov.uk/economy/nationalaccounts/balanceofpayments/bulletins/uktrade/january2023#:~:text=The%20trade%20in%20goods%20deficit,billion%20to%20%C2%A336.5%20billion.

9. "Read Donald Trump's Speech on Trade," *Time*, June 28, 2016, https://time.com/4386335/donald-trump-trade-speech-transcript/#lnkibaj38by47dfvnqd.

10. JaeBin Ahn, Benjamin Carton, Ashique Habib, Davide Malacrino, Dirk Muir, and Andrea Presbitero, "Geoeconomic Fragmentation and Foreign Direct Investment," in *World Economic Outlook, April 2023: A Rocky Recovery* (Washington: International Monetary Fund, 2023).

and leaving low-income ones out. He pointed to two US trade agreements with the European Union and Japan focused on trade in batteries and critical minerals.[11]

Martin Stuermer commented on electric car batteries being imported from countries considered not friendly to the United States. Stuermer asked the authors for their experience evaluating the impacts of local content requirements and if they could predict these policies' impact on future trade in batteries as well as critical minerals.

Goldberg clarified what local content requirements are. She explained that American firms producing semiconductors are not currently much affected by export restrictions. However, these new policies require any country utilizing US technology to gain approval to export to China. She added that this inevitably forces countries to choose between the United States and China.

Steven Davis questioned whether geopolitical tension and conflict have a negative impact on economic growth and innovation. Davis cited the US Defense Department's investments in new technologies during the Cold War and the US space program, both spurred by geopolitical rivalry with the Soviet Union. He also referred to China's recent investments in leading technologies and argued that some were made in what China perceived as its national security interests. His last example was the push to develop and mass-produce new vaccines after the onset of the COVID-19 pandemic. That push also had overtones of geopolitical competition. It's possible, Davis suggested, that the positive effects of geopolitical tensions on innovation and research outweigh the negative impacts on trade.

Goldberg responded to Davis's comments and agreed that geopolitical tension leads to investments in innovation. She questioned whether there were better investments that could have been made. She suggested that there are less harmful reasons to invest in innovation.

Reed suggested there are also channels through which geopolitical tensions could harm innovation. He cited recent research that found the US Justice Department's investigations on Chinese researchers in the United States were followed by a decline in publications by and citations of scientists with previous collaborations with scientists in China, even if these

11. International Trade Administration, "EU—Country Commercial Guide," https://www.trade.gov/country-commercial-guides/eu-trade-agreements; Office of the United States Trade Representative, "U.S.-Japan Trade Agreement Text," https://ustr.gov/countries-regions/japan-korea-apec/japan/us-japan-trade-agreement-negotiations/us-japan-trade-agreement-text.

scientists were not themselves subject to investigations.[12] Reed added that regardless of the effects of innovation, increased trade restrictions would make it more difficult to collaborate and create low-cost, high-scale technology. He used solar panels as an example of a technology developed in the United States that is now cheaper and more prevalent because of production capacity abroad.

Donald Kohn brought up the intersection of globalization and inflation. He compared globalization to a positive supply shock and deglobalization to a negative supply shock. Kohn pointed out the figures in Goldberg and Reed's paper that showed globalization's peak in 2009–2010 followed by a period of decline. He noted that inflation was low during the initial onset of deglobalization and questioned whether globalization has a strong impact on inflation.[13]

Jonathan Pingle built upon Kohn's point and referred to the period following World War I. He noted that the postwar period was a time of both deglobalization and disinflation, even amid immigration restrictions. He cited the paper's evidence of both FDI flows and the increase in international trade immediately following the pandemic. Rather than assume deglobalization will occur, he suggested questioning whether it will and how impactful the effects would be.

Goldberg pointed out the distinction between the level and direction of globalization. She noted that although the world economy may be moving in the direction of less globalization, globalization itself still exists. She commented on the unprecedented monetary policy actions that took place after the 2008 financial crisis that call for an explanation of why inflation remained low until 2021. Goldberg suggested that, although there is no clear econometric evidence on this phenomenon, globalization likely played a role in controlling inflation. To support her point, she referred to globalization's impacts on workers' bargaining power. She stated that if workers know their job can be outsourced to a low-wage country or performed by technology, they are less able to negotiate for a higher salary. She attributed some of the low levels of inflation to this loss in bargaining power.

Goldberg continued with a discussion of labor and the global integration of labor markets. She argued that trade differs from immigration in two

12. Ruixue Jia, Margaret E. Roberts, Ye Wang, and Eddie Yang, "The Impact of US-China Tensions on US Science," working paper 29941 (Cambridge, Mass.: National Bureau of Economic Research, 2022).

13. International Monetary Fund, "Inflation Rate, Average Comsumer Prices," https://www.imf.org/external/datamapper/PCPIPCH@WEO/OEMDC/ADVEC/WEOWORLD.

major ways. She contrasted the bipartisan support for domestic production with the polarization produced by immigration policy. Additionally, she argued that, as the United States continues to distance itself from low-cost trade partners, keeping labor laws flexible and allowing migration can help reduce inflationary pressure.

Reed closed the discussion by noting the impact Russia's invasion of Ukraine had on public sentiment toward deglobalization. He referred to the book *War by Other Means*, which argues for economists having less authority in foreign policy because their estimations can differ from national security experts.[14] He added that this changing sentiment has increased over time and became especially salient after the Russian invasion of Ukraine.

14. Robert D. Blackwill and Jennifer M. Harris, *War by Other Means: Geoeconomics and Statecraft* (Cambridge, Mass.: Harvard University Press, 2016).

AMY FINKELSTEIN
Massachusetts Institute of Technology

OWEN ZIDAR
Princeton University

CASEY McQUILLAN
Princeton University

ERIC ZWICK
University of Chicago Booth School of Business

The Health Wedge and Labor Market Inequality

ABSTRACT Over half of the US population receives health insurance through an employer with premium contributions creating a flat "head tax" per worker, independent of their earnings. This paper develops and calibrates a stylized model of the labor market to explore how this uniquely American approach to financing health insurance contributes to labor market inequality. We consider a partial-equilibrium counterfactual in which employer-provided health insurance is instead financed by a statutory payroll tax on firms. We find that, under this counterfactual financing, in 2019 the college wage premium would have been 11 percent lower, noncollege annual earnings would have been $1,700 (3 percent) higher, and noncollege employment would have been nearly 500,000 higher. These calibrated labor market effects of switching from head tax to payroll tax financing are in the same ballpark as estimates of the impact of other leading drivers of labor market inequality, including changes in outsourcing, robot adoption, rising trade, unionization, and the real minimum wage. We also consider a separate partial-equilibrium counterfactual in which the current head tax financing is maintained, but 2019 US health care spending as a share of GDP is reduced to the Canadian share; here, we estimate that the 2019 college wage premium would have been 5 percent lower and noncollege

Conflict of Interest Disclosure: This research was supported by the Laura and John Arnold Foundation. Owen Zidar was also supported by the National Science Foundation, grant number 1752431, and Eric Zwick was supported by the Fama Research Fund, Neubauer Family Foundation, the Polsky Center, and the Hultquist Faculty Research Endowment at Chicago Booth. Other than the aforementioned, 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. The authors are not currently an officer, director, or board member of any organization with a financial or political interest in this paper.

Brookings Papers on Economic Activity, Spring 2023: 425–475 © 2024 The Brookings Institution.

annual earnings would have been 5 percent higher. These findings suggest that health care costs and the financing of health insurance warrant greater attention in both public policy and research on US labor market inequality.

The gap in labor market outcomes between college- and noncollege-educated workers has widened in the United States over the last four decades. In 2019, the wages of college-educated workers were nearly twice as high as noncollege-educated workers, and college-educated workers also had much higher employment rates. A large and storied body of literature has explored the causes of this labor market inequality and its spectacular rise. This literature has uncovered a variety of contributing factors, including skill-biased technological change (Autor, Goldin, and Katz 2020; Katz and Murphy 1992; Bound and Johnson 1992; Goldin and Katz 2008; Acemoglu and Autor 2001), institutional changes such as the erosion of unions and worker bargaining power and a declining real minimum wage (Card, Lemieux, and Riddell 2004; Farber and others 2021; DiNardo, Fortin, and Lemieux 1996; Lee 1999), globalization (Feenstra and Hanson 2003; Goldschmidt and Schmieder 2017), and the sorting of workers across firms (Abowd, Kramarz, and Margolis 1999; Card, Heining, and Kline 2013; Song and others 2019). The increase in labor market inequality has not been limited to the United States, and many of these same forces may also drive similar trends in other Organisation for Economic Co-operation and Development (OECD) countries. Yet the level and growth in labor market inequality are particularly pronounced in the American case.

A uniquely American factor that may contribute to labor market inequality is the financing of health insurance through the workplace. About half of the US population—and virtually all of those with private health insurance—receive their health insurance through their employer or a family member's employer.[1] The government heavily subsidizes employer-provided health insurance by excluding any contribution employers make to their employees' health insurance premiums from employees' taxable income. This tax exclusion is the single largest federal tax expenditure. It costs the federal government about $300 billion a year (Congressional Budget Office 2019), or about two-fifths of the amount it spends on Medicare.[2]

1. Kaiser Family Foundation, "Health Insurance Coverage of the Total Population" [data set], www.kff.org/other/state-indicator/total-population/.

2. US Centers for Medicare and Medicaid Services, "National Health Expenditure Accounts: Historical," https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.

A lot of literature in public finance has analyzed the impact of this tax subsidy on health insurance coverage, health care spending, and (skill-neutral) labor market distortions brought about by so-called job-lock (Feldstein 1973; Feldstein and Friedman 1977; Gruber 2000, 2002; Garthwaite, Gross, and Notowidigdo 2014; Gruber and Madrian 2004). However, the potential role for firm-based health care financing to contribute to labor market inequality in the United States has received comparatively little attention in either the public finance or the inequality literature.

Crucially, from the perspective of labor market inequality, health care costs for workers do not decrease as their earnings fall. Therefore, unlike other employee benefits that are designed to replace lost earnings, such as life insurance, disability insurance, or unemployment insurance, the cost of providing a worker with a given health insurance plan is a fixed dollar cost per worker, regardless of wage or earnings. This increases the price of lower-skilled labor relative to higher-skilled labor, a phenomenon we refer to as the "health wedge." The health wedge is substantial. Average insurance premiums for employer-provided health insurance were about $12,000 in 2019. This amount is about 25 percent of the average annual earnings for a full-time, full-year worker without a college education (about $50,000), and about 12 percent of the average annual earnings for a full-time, full-year college-educated worker (about $100,000; see table 1, panel A).

Several leading economists have recently and prominently conjectured that the health wedge has influenced US labor market inequality. Emmanuel Saez and Gabriel Zucman advance this hypothesis in *The Triumph of Injustice* (2019a), and summarize it in the popular press:

> Because health insurance premiums are fixed, the wage penalty is the same for a low-wage secretary as it is for a highly paid executive. This severely depresses wages for tens of millions of moderate-income workers. . . . It's the most unfair type of tax: A huge burden for low-wage workers and almost meaningless for the rich. (Saez and Zucman 2019b, pars. 6, 8)

Anne Case and Angus Deaton make a similar argument in *Deaths of Despair and the Future of Capitalism* (2020b), which they summarize in an op-ed:

> Employer-based health insurance is a wrecking ball, destroying the labor market for less-educated workers. . . . At the very least, America must stop financing health care through employer-based insurance, which encourages some people to work but it eliminates jobs for less-skilled workers. (Case and Deaton 2020a, pars. 11, 23)

This qualitative observation follows naturally from textbook models. The data are also suggestive of a potentially important role for the health wedge in contributing to US labor market inequality. Between 1977 and 2019,

both the health share of GDP and the college wage premium roughly doubled (figure 1, panel A). Relative to other large OECD countries, the United States today is an outlier both in terms of the size of its health care sector and the college wage premium (figure 1, panel B). However, to our knowledge, we have little evidence of the quantitative importance of the health wedge for labor market inequality.

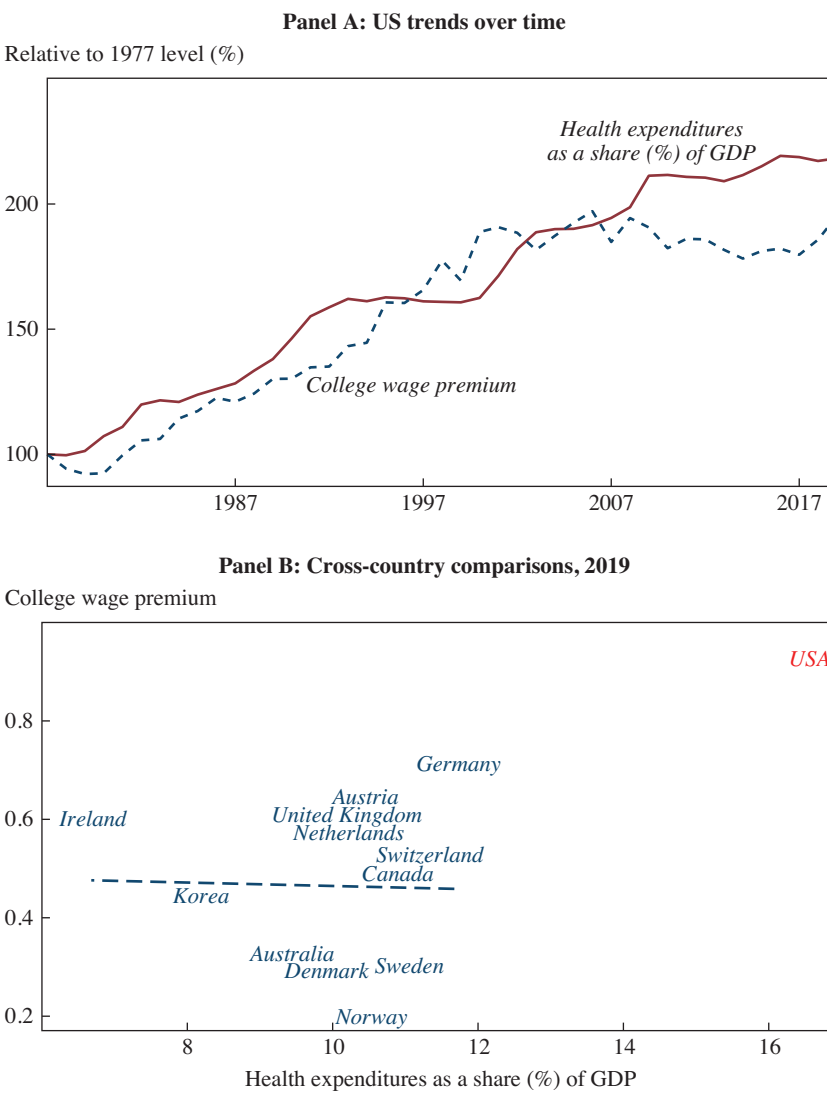
In this paper, we develop and calibrate a simple model of the labor market and use it to explore quantitatively how the US approach to health insurance financing may contribute to labor market inequality. Specifically, we ask what labor market outcomes would have been for full-time, full-year workers under two types of partial-equilibrium counterfactuals.

The first set of counterfactuals considers an alternative financing of health insurance through a national payroll tax on firms rather than through the current head tax approach. Specifically, we calculate what labor market outcomes would have been if employees who receive health insurance through employer-provided head tax financing instead received it financed through a national payroll tax proportional to earnings that is levied on firms. This payroll tax financing approach is similar in spirit to how universal health insurance is financed in many countries, such as Canada and Germany. Our purpose is not to propose such a change in financing per se, but rather to use a realistic counterfactual financing approach to quantify the impact of the current head tax financing on labor market inequality. To focus on the impact of a change in financing of a given amount of insurance coverage, we do not give firms the option to stop or start offering health insurance. We also abstract from other potential margins of firm adjustment (such as their decisions around part-time work, outsourcing, or offshoring).

We find that, under this counterfactual payroll tax financing, the college wage premium would have been about 11 percent lower in 2019, noncollege annual earnings would have been approximately $1,700 (about 3 percent) higher, and noncollege employment would have been nearly 500,000 higher. Had this counterfactual financing been in place since 1977, the rise in the college wage premium would have been about 20 percent smaller and the rise in noncollege employment about 4.6 percent larger.[3]

3. Naturally these magnitudes depend on the assumptions we make about empirical objects such as the size of the head tax, the differences in productivity across skill groups, and labor supply functions. We explore the sensitivity of the calibration results to alternative assumptions for these parameters and find the results reassuring. For example, a range of alternative parameters suggests that, under payroll tax financing, the 2019 college premium would have been 10 to 13 percent lower (compared to our baseline estimate of 11 percent), and the rise in the college wage premium between 1977 and 2019 would have been 18 to 23 percent smaller (compared to our baseline estimate of 20 percent).

Figure 1. College Wage Premiums and Health Expenditures



Sources: Authors' calculations using Current Population Survey and OECD data.

Note: Panel A shows health expenditures as a share of GDP and the college wage premium. Both are measured as a percentage relative to their 1977 level. Data on college wage premiums are from the Current Population Survey and are defined as the ratio of full-time, full-year earnings for college workers age 25–64 relative to noncollege workers, minus one. Data describing health care as a share of GDP are from OECD Global Health Expenditure. Panel B plots the relationship between the college wage premium and health expenditures in 2019 for countries with a GDP above 2019 US$ 300 billion. The dashed line shows the best-of-fit for non-US countries, weighted equally. The college wage premium is defined as $W_C/W_N - 1$, or the ratio of college to noncollege wages minus one. Non-US data for twelve countries are from the OECD, further described in online appendix A.1.

These effects are comparable in magnitude to estimates of some of the other leading drivers of labor market inequality, including outsourcing, robot adoption, rising trade, declining unionization, and the decline in the real minimum wage.

Our second set of counterfactuals considers how the spectacular rise in US health care spending and health insurance premiums over the last four decades has affected labor market inequality under the current, head tax regime. Between 1977 and 2019, average health insurance premiums for employer-provided health insurance rose by about $9,000 (in 2019 dollars). This rise was largely in response to the substantial rise in health care spending. We find that if, counterfactually, US health care spending as a share of GDP in 2019 had remained at the 1977 level of 7.7 percent of GDP rather than its 2019 level of 16.8 percent, the college wage premium would be about 11 percent lower and noncollege wages would be about $6,000 (12 percent) higher. In a somewhat more realistic counterfactual, we find that if US health care spending as a share of GDP in 2019 had been the same as the Canadian share—that is, approximately 10.8 percent instead of 16.8 percent of GDP—the college wage premium would have been 5 percent lower and noncollege annual earnings would have been $2,800 (5 percent) higher.

Our analyses rely on several simplifying assumptions. Perhaps most importantly, they occur in partial equilibrium and therefore do not provide a full, general equilibrium assessment of the potential impact of a change in health care financing or spending. Among other things, in all of our analyses we hold constant the share of full-time, full-year workers covered by health insurance as well as the comprehensiveness of employer-provided health insurance coverage. These could be affected by our counterfactuals; indeed, related literature on the labor market impacts of other health insurance reforms endogenizes some of these factors, such as the decision of firms to offer health insurance and the decision of workers to sort into firms with or without health insurance (Dey and Flinn 2005; Aizawa 2019; Aizawa and Fang 2020; Fang and Krueger 2021).[4] Relatedly, we abstract from the ways in which the head tax financing of employer-provided health insurance might also contribute to the "hollowing out" of the workforce (Autor 2018; Autor and Dorn 2013), a shift to part-time workers (Cutler

4. In the most closely related work that we know of, Beemon (2021) uses a labor search model to estimate the impact of switching from employer-provided health insurance to free public insurance on the equilibrium distribution of wages, finding that this would reduce wage inequality.

and Madrian 1998), the rise of alternative work arrangements (Katz and Krueger 2017), and the fissuring of the workforce (Weil 2014; Card, Heining, and Kline 2013; Song and others 2019). Nonetheless, our stylized, partial equilibrium analysis points to the potential importance of this uniquely American form of health care financing in contributing to labor market inequality. Our findings suggest that the financing of US health insurance warrants greater attention in both public policy and research on US labor market inequality.

The rest of the paper proceeds as follows. Section I provides background on employer-provided health insurance and on patterns of labor market inequality. Section II describes a simple model of the effects of health insurance financing on labor market inequality. Section III discusses our calibration. Section IV presents the main results on labor market outcomes under counterfactual payroll tax financing. Section V examines labor market outcomes under counterfactual levels of health care spending and health insurance premiums. Section VI concludes.

I. Background

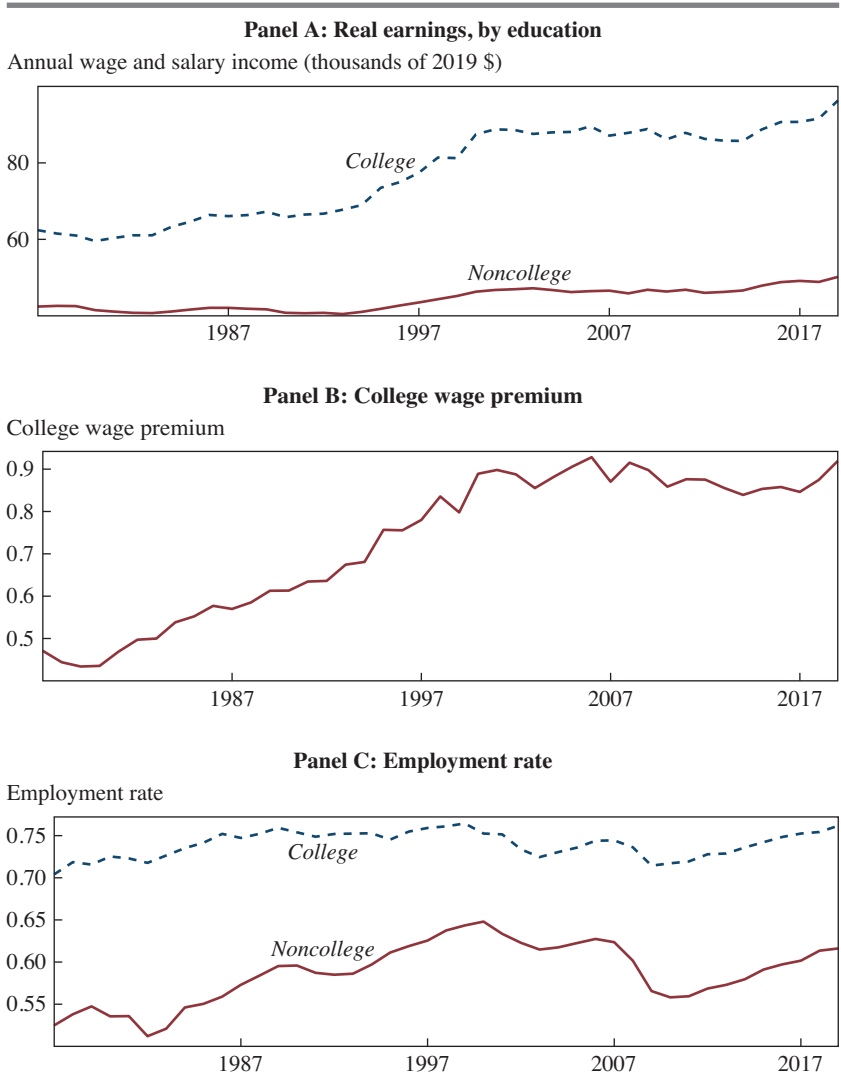
I.A. Trends in Labor Market Inequality and Health Care Spending

Labor market inequality has risen dramatically over the past few decades in the United States. Figure 2 shows trends in labor market outcomes for full-time, full-year workers age 25–64 from the Current Population Survey (Flood and others 2021). Full-time workers are defined as people who worked at least forty weeks in the year and had a usual work week of at least thirty hours. We focus on full-time, full-year workers to simplify the measurement of wages and because employer-provided health insurance is much more common among this group. We report trends separately for those with a college degree—defined as a bachelor's degree or higher—and those without a college degree.

Real annual earnings for college-educated workers (w_C) rose from about \$63,000 (in 2019 dollars) in 1977 to nearly \$100,000 in 2019.[5] At the same time, real annual earnings for noncollege workers (w_N) grew more gradually, from about \$43,000 in 1977 to about \$50,000 in 2019. As a result, the college

5. The rise in wages for college-educated workers was smaller in the 2000s, which could be due in part to the rise of business income of entrepreneurial owner-managers (Smith and others 2019) who face tax incentives to re-characterize wages as profits and whose income is large enough to affect aggregate trends in the corporate sector labor share (Smith and others 2022). Throughout, all dollar values are adjusted to 2019 US dollars using the FRED series Personal Consumption Expenditures (PCE) price index.

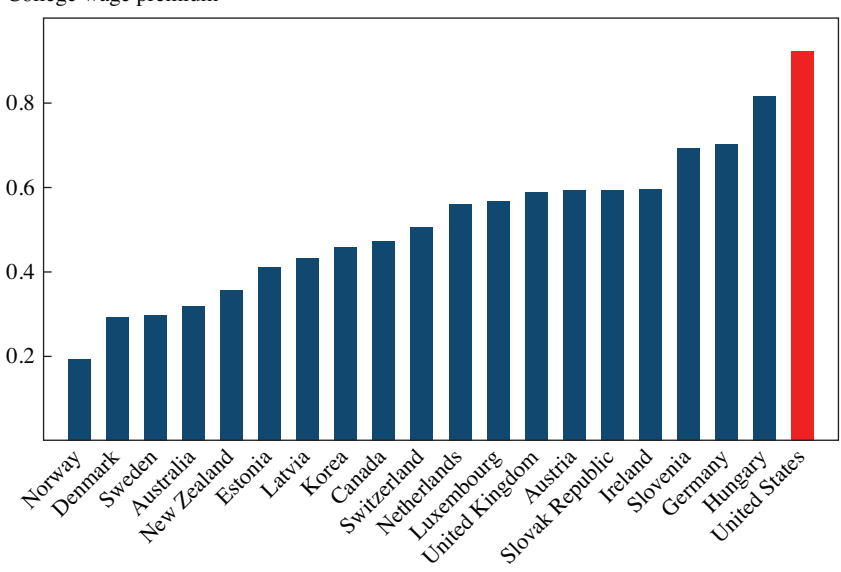
Figure 2. Labor Market Outcomes by Education



Source: Authors' calculations using the Current Population Survey.

Note: Panel A shows the average real wages of college- and noncollege-educated workers. Panel B shows the college wage premium, defined as the ratio of college to noncollege wages minus one. Panel C shows the employment rate, defined as the ratio of workers to the population, for college- and noncollege-educated individuals. The population is restricted to individuals age 25–64. Workers are defined as those employed full-year (at least forty weeks per year) and full-time (at least thirty hours a week). Wages are calculated in 2019 dollars using a PCE price index. Using the CPI price index instead, noncollege workers' wages fall $1,779 instead of increasing by $7,754, and college workers' wages increase $19,880 instead of by $33,903; however, the choice of price index has no effect on the evolution of the college wage premium or the employment rate. Based on guidance from the editors, we use the PCE price index as the baseline.

Figure 3. College Wage Premiums across Countries in 2019



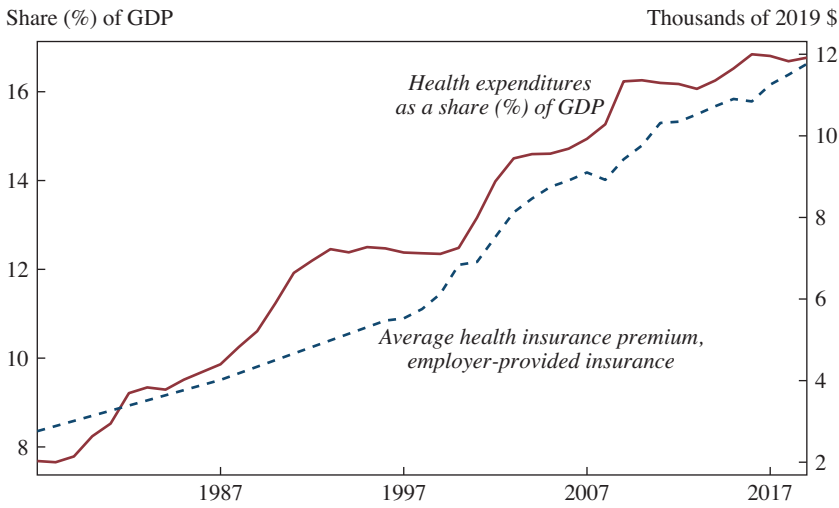
Source: Authors' calculations using data from OECD Education and Earnings, https://stats.oecd.org/Index.aspx?DataSetCode=EAG_EARNINGS.

Note: This figure plots the ratio of college-educated to noncollege-educated worker wages by country. College wage premiums are given for the population of those age 25–64 who were full-year, full-time workers in 2019. The OECD defines full-time workers as those working at least thirty hours per week and full-year according to each country's individual definition.

wage premium $\left(\frac{w_C}{w_N} - 1\right)$ increased from 47 percent in 1977 to 90 percent in 2019. This 90 percent college wage premium exceeds that of other countries for which comparable data are available, including the United Kingdom (figure 3).

Over the same time period, real premiums for employer-provided health insurance in the United States have also risen substantially, fueled by rising health care spending (figure 4). Since 1977, the average health insurance premium has quadrupled in real terms, from about $2,750 (in 2019 dollars) to about $12,000 in 2019. At the same time, health care spending as a share of GDP has risen from 7.7 to 16.8 percent. As a result of these trends in health insurance premiums and in earnings, health insurance premiums as a fraction of labor market earnings increased between 1977 and 2019 from 6 percent of noncollege earnings to almost 25 percent, and from 4 percent of college earnings to 12 percent.

Figure 4. US Health Expenditures and Insurance Premiums



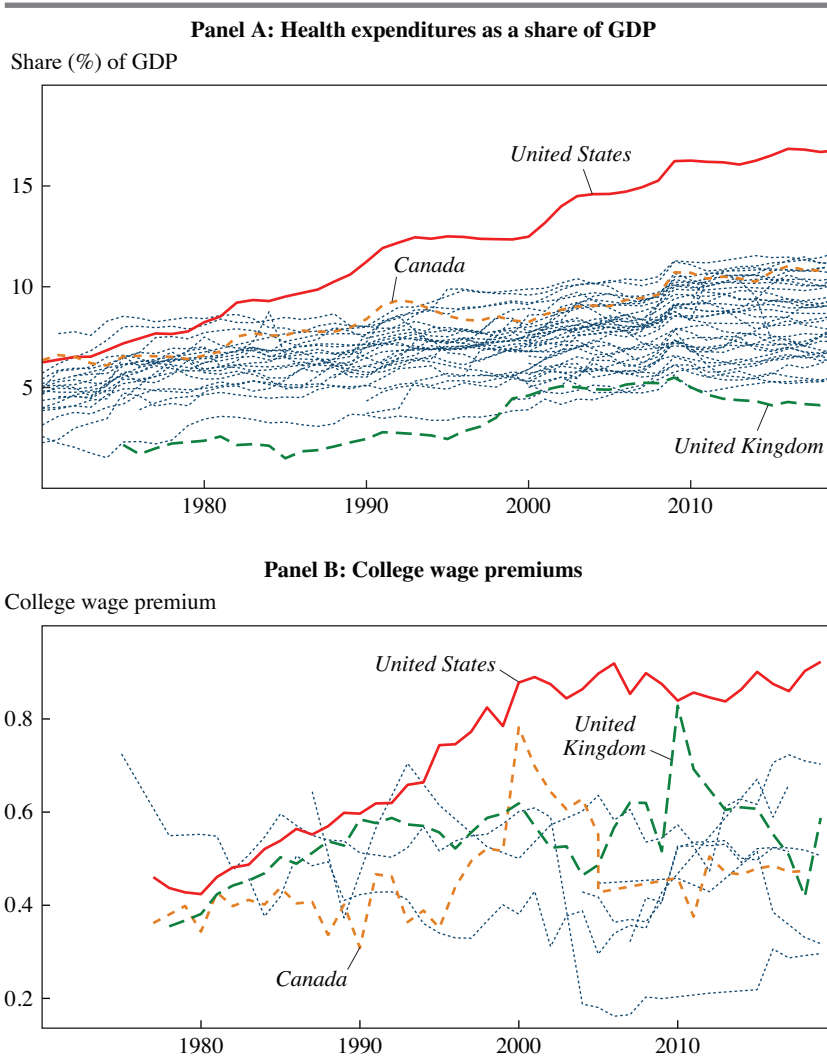
Sources: OECD; National Medical Care Expenditure Survey (1977), https://www.icpsr.umich.edu/web/NACDA/studies/8325; National Medical Expenditure Survey (1987), https://www.icpsr.umich.edu/web/ICPSR/studies/9674; and Medical Expenditure Panel Survey (1996–2019), https://meps.ahrq.gov/data_stats/download_data_files.jsp.

Although many OECD countries experienced both an increase in health care spending as a share of GDP and an increase in labor market inequality, the United States is an outlier in both trends (figure 5). Health care spending as a share of GDP in the United States rose from about 8.2 percent of GDP in 1980 to 16.8 percent in 2019. At the same time, on average across the OECD countries (excluding the United States) it rose from 6 to 9 percent (OECD 2019).[6] Likewise, the college wage premium in the United States increased from 42 percent in 1980 to 91 percent in 2019. Over the same period, in the United Kingdom, Sweden, and Canada, the college premium rose on average from 43 percent in 1980 to only 45 percent in 2019 (Brzozowski and others 2010; Domeij and Flodén 2010; Blundell and Etheridge 2010; Eurostat 2022; OECD 2022).[7]

6. Across the United Kingdom, Sweden, and Canada, where we can also calculate the change in the college wage premium, health care spending as a share of GDP rose on average from 6.5 percent in 1980 to 10.6 percent in 2019.

7. Estimating the college premium for all OECD countries over this period is beyond the scope of this paper. That said, cross-country evidence (e.g., Krueger and others 2010) suggests that most other OECD countries experienced much smaller increases in the college premium over the past few decades. Despite having college premiums closer to the United States in 1980, most OECD countries have lower college premiums than the United States does today (figure 3).

Figure 5. Health Expenditures and College Wage Premiums across Countries



Sources: Brzozowski and others (2010); Eurostat (2022); Fuchs-Schuendeln, Krueger, and Sommer (2010); and Jappelli and Pistaferri (2010).

Note: The college wage premium is defined as $W_C/W_N - 1$. In both panels, the United States, the United Kingdom, and Canada are labeled. Panel A includes all other OECD countries. Additional countries in panel B are Australia, France, Germany, Italy, Sweden, and Switzerland. A version of panel B with a full set of OECD countries can be found in figure A.2 in the online appendix; data sources and construction are described in online appendix A.1.

These patterns over time and across countries lend some credence to the hypothesis that US health care—and in particular the financing of health insurance through employers—may be contributing to rising labor market inequality in the United States.

I.B. Employer-Provided Health Insurance

INSTITUTIONAL BACKGROUND The workplace is the primary source of private health insurance in the United States. About half of the US population—and virtually all of those with private insurance—receive that insurance through an employer.[8] This development is generally viewed as an accident of history. During World War II, the federal government imposed wage and price controls on American firms as part of its effort to prevent a surge in inflation in the face of competition for scarce labor and goods. But employer contributions to health insurance didn't "count" as part of workers' wages, and employers soon realized that this created a loophole: faced with binding wage controls, they started offering—and paying for—workers' health insurance as a way to attract and retain employees. What had been initially viewed as a wartime stopgap measure became codified and entrenched into the tax code after the war, with the 1954 codification of the exclusion of employer contributions to health insurance from taxable income. It remains in place to this day (Starr 1982; Thomasson 2002). Thomasson (2003) argues that this 1954 codification contributed to the rise of employer-provided health insurance in the United States.

While employer compensation paid in the form of wages and salary is subject to personal income taxes and to payroll taxes on both the employee and employer, compensation paid in the form of contributions to health insurance premiums is not. This treatment creates a tax subsidy to employer-financed health insurance s that is given by:

$$s = 1 - \left(\frac{1 - \tau_{inc} - \tau_{ss}}{1 + \tau_{inc}} \right),$$

where τ_{inc} is the employee's marginal tax rate on earnings and τ_{ss} is the statutory payroll tax rate paid by the employee and separately by the employer. To see where this formula comes from, first note that the employer is indifferent between contributing a dollar to the employee's health insurance

8. Kaiser Family Foundation, "Health Insurance Coverage of the Total Population" [data set], www.kff.org/other/state-indicator/total-population/.

premiums and contributing $1/(1 + \tau_{ss})$ to wages, since the employer must pay payroll tax on any wage contributions but not on health insurance contributions; we assume the incidence of payroll taxes is fully on the worker. If employees are paid $1/(1 + \tau_{ss})$ in wages, they must in turn pay both income tax and payroll tax on that wage. Thus, workers face a choice of receiving \$1 in employer contributions to their health insurance premiums or $\left(\frac{1 - \tau_{inc} - \tau_{ss}}{1 + \tau_{ss}}\right)$ in take-home pay.

The resulting tax subsidy is substantial. For example, in 2019 the combined employer and employee payroll tax rate—including both Social Security and Medicare—was 15.3 percent, split evenly between employer and employee ($\tau_{ss} = 7.65$).[9] An employee earning \$50,000 in 2019 faced a federal income tax rate of 22 percent while one earning \$100,000 faced a 24 percent tax rate (El-Sibaie 2018). Thus, the tax subsidy to employer-financed health insurance s was about 35 percent; if the employee faced state income taxes, it would be even higher. In other words, at the same cost to the employer, the employee could receive either \$1 in contributions to health insurance premiums or \$0.65 in take-home pay.

The tax subsidy to employer-provided health insurance is uniformly reviled by economists of both parties (Mankiw and Summers 2015).[10] A large body of empirical literature reviewed by Gruber (2002) has documented that the tax subsidy distorts compensation from wages to health insurance; this in turn distorts the demand for medical care (Feldstein 1973; Feldstein and Friedman 1977; Pauly 1986). The tax subsidy is also highly regressive, since both rates of employer provision of health insurance and tax rates rise with employee income (Pauly 1986; Gruber 2011). There have been policy attempts to reduce the tax subsidy, most notably the so-called Cadillac tax under the 2010 Affordable Care Act, which would have reduced the tax subsidy to employer-provided health insurance premiums above a specified dollar amount. It was passed into law but never put into effect.[11]

9. The Social Security component of the tax rate was not applied to income above \$132,900; Social Security, "Contribution and Benefit Base," https://www.ssa.gov/oact/cola/cbb.html.

10. See also Chicago Booth Kent A. Clark Center for Global Markets, "U.S. Poll: Cadillac Tax," https://www.kentclarkcenter.org/surveys/cadillac-tax/.

11. There have been many related policy proposals. For example, concerned about the tax subsidy's regressivity and its encouragement of low-value health care innovation, Bagley, Chandra, and Frakt (2015, 11) suggest that Congress should "replace the tax exclusion for health insurance with a tax credit for employer-sponsored insurance—a fixed amount that each taxpayer could subtract from her overall tax liability—that phases out as income increases. Less radically, the tax exclusion could itself phase out with income."

As to the provision of health insurance through employers more broadly, the prevailing wisdom is that it is hard to rationalize on efficiency grounds. "If we had to do it over again," the health economist Uwe Reinhardt observed, "no policy analyst would recommend this model" (Blumenthal 2006, 82). A large body of empirical literature has documented that the linking of insurance to the workplace distorts labor market decisions, including retirement and labor supply, and limits job-to-job mobility (Gruber 2000; Gruber and Madrian 2004; Garthwaite, Gross, and Notowidigdo 2014). More closely related to our head tax analysis, Cutler and Madrian (1998) observe that because health insurance contributions represent a fixed cost per employee, rising health care costs should encourage a reduction in the number of employees and an increase in hours per worker. Consistent with this prediction, they find that hours per worker increased more over time for workers with employer-provided health insurance compared to those without.

The head tax feature of financing employer-provided health insurance also raises the question of whether firms can voluntarily use a financing approach that charges highly compensated employees a larger amount for health insurance than lower-paid employees. Our understanding is that firms can do so under current law but rarely do. Nondiscrimination rules under IRS Code section 105(h) prohibit discrimination in favor of highly compensated individuals, however no regulations forbid the opposite. Nevertheless, it appears that in practice market forces largely prevent firms from pursuing this approach.[12] Charging highly compensated workers more for health benefits effectively lowers their wage, making this approach less attractive for firms that compete to retain and attract college workers. Offsetting a higher charge for benefits with more pay would be a disadvantage tax-wise relative to the status quo arrangement.

In the context of cost sharing, Robertson (2014) argues that agency frictions and the locus of benefit decisions within the firm can also explain why firms are reluctant to implement progressive financing arrangements in their health plans. Specifically, the managers who make such decisions tend to be the higher-income workers most likely to be hurt by a proportional-to-income scheme. Customizing plans to depend on worker income at the firm level also introduces administrative burden in designing health benefit

12. Recent surveys suggest that a number of large employers do vary contributions by coarse salary classes, but this is a far cry from the proportional financing we will analyze below (Gregware 2017; Sammer 2017). Such arrangements are likely driven by the affordability limit under the Affordable Care Act, which requires the lowest-cost single coverage plan offered by large employers to be below 9.5 percent of household income.

Table 1. Summary Statistics for Full-Time, Full-Year Workers Age 25–64 (2019)

	Total	*College*	*Noncollege*
Panel A: Labor market outcomes			
FTFY employment rate (P_g)	0.672	0.762	0.616
Average annual earnings (w_g)	$70,333	$96,304	$50,179
Panel B: Health insurance coverage			
Employer-sponsored	0.802	0.895	0.729
Policyholder	0.659	0.732	0.603
Dependent	0.140	0.162	0.123
Other private	0.062	0.055	0.067
Public	0.072	0.032	0.103
None	0.084	0.031	0.125
Panel C: Offering and take-up			
Offered employer-sponsored health insurance	0.830	0.895	0.780
Take-up | Offered	0.794	0.818	0.773

Source: Current Population Survey.

Note: College-educated workers are defined as those who have obtained a bachelor's degree or above; noncollege-educated workers do not have a bachelor's degree. The full-time, full-year (FTFY) employment rate is the share of the population who worked at least forty weeks in the year and had a usual work week of at least thirty hours. Offering is defined as individuals who either enroll in employer-provided health insurance as a policyholder or report being offered this insurance at their workplace; take-up is defined as enrolling in employer-provided health insurance as a policyholder. The results are similar when looking at self-reported eligibility for employer-provided health insurance instead of offering. Dollar amounts are in 2019 US dollars.

menus relative to the status quo. Consistent with these barriers to private implementation, conversations with employment law experts suggest that firms typically approach health benefits from the perspective of how much they will cost to provide and then offer a simple fixed price per worker in each coverage class.

DESCRIPTIVE STATISTICS Table 1 provides an overview of the health insurance of the approximately 100 million full-time, full-year workers age 25–64, based on the 2019 Current Population Survey. Panel A summarizes the key labor market outcomes in 2019 that will be the focus of our analysis: the employment rate and average annual earnings. The first row indicates that two-thirds of those age 25–64 were working full-time, full-year in 2019, with the rest either working less than full-time, full-year or not working. The full-time, full-year employment rate for college-educated workers is 0.76 and for noncollege-educated workers it is 0.62. Average annual earnings are $96,304 for full-time, full-year college-educated workers (w_C) and $50,179 for full-time, full-year noncollege-educated workers (w_N). Figure 2, panel C, shows that the full-time, full-year employment rate for college-educated workers (P_C) has risen since 1977 from about 0.70 to 0.76, while that for noncollege-educated workers (P_N) has risen from about 0.52 to 0.62.

We focus on the extensive margin of labor force participation; rates of labor force nonparticipation are high among noncollege-educated prime-age adults, even for men, and have generated substantial interest in their causes and consequences (Binder and Bound 2019).

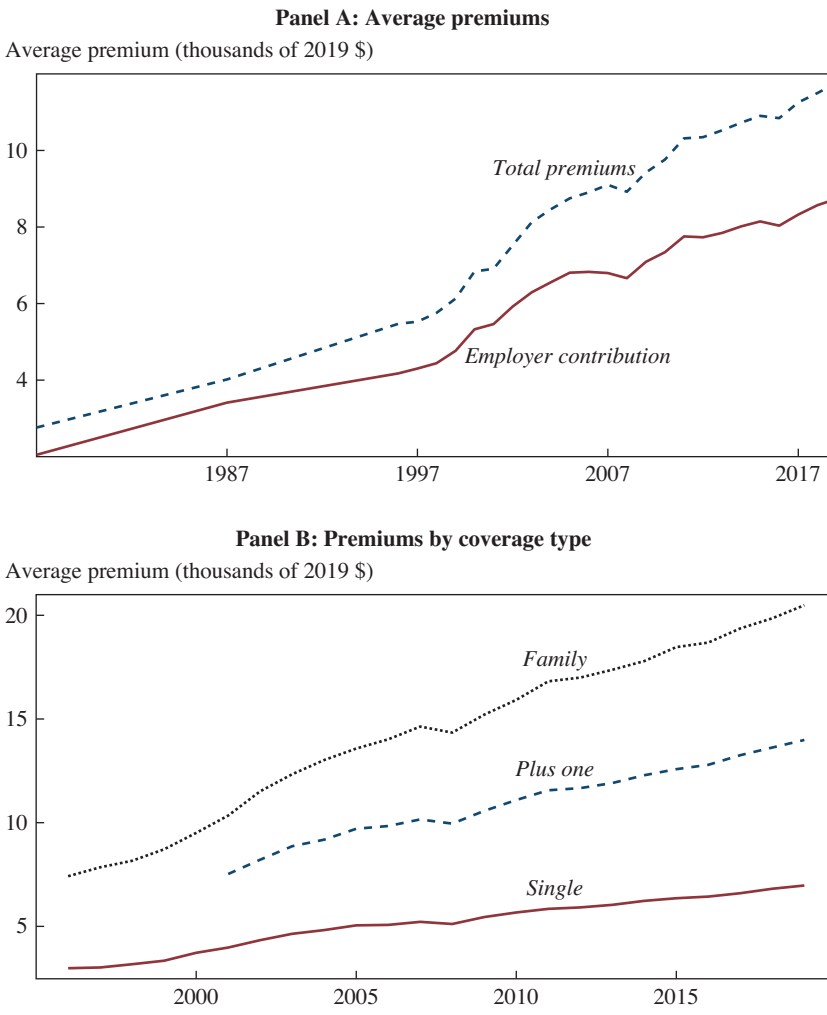
Table 1, panel B, describes the health insurance coverage of these full-time, full-year workers. Just over 80 percent are covered by employer-provided health insurance, 13 percent have another form of health insurance (e.g., non-group private health insurance or public insurance such as Medicaid), and 8 percent are uninsured. About two-thirds of full-time workers are policyholders of employer-provided health insurance—meaning that any employer contributions to those premiums are part of the cost of hiring such workers—while another 14 percent have coverage as a dependent on a spouse's employer-provided health insurance. The share of workers who are policyholders for employer-provided health insurance is higher for college-educated workers (73 percent) than noncollege-educated workers (60 percent) while rates of uninsurance are lower (3 percent versus 12.5 percent).[13]

The one-third of workers who are not policyholders reflects a combination of working for a firm that doesn't offer health insurance and not taking up offered insurance, in roughly equal measure (table 1, panel C). About 83 percent of full-time, full-year workers are offered insurance by their employer, and, conditional on being offered this insurance, about 79 percent take up this insurance (i.e., enroll as the policyholder). Offering and conditional take-up are higher for college-educated workers (at 90 percent and 82 percent, respectively) than for noncollege-educated workers (78 percent and 77 percent, respectively). One reason that workers may not take up employer-provided health insurance is that they typically have to pay a portion of the premiums; many who do not take up the offered insurance through their employer are insured through another source, such as another family member's employer-provided health insurance or public insurance (Gruber 2008; Gruber and Simon 2008).

Average annual premiums for employer-provided health insurance have been rising over time (figure 6, panel A). The data on premiums are provided directly by employers and include average employer and employee

13. Another 14 percent of those age 25–64 are part-time workers, defined as anyone who reports working (not counting self-employment) during the year but does not meet the definition of a full-time, full-year worker. Only 53.5 percent of them are covered by employer-provided health insurance, and only 31 percent are policyholders; 15 percent have no health insurance.

Figure 6. Premiums for Employer-Provided Health Insurance



Sources: Medical Expenditure Panel Survey (1996–2019); National Medical Expenditure Survey (1987); and National Medical Care Expenditure Survey (1977).

Note: The average total premium is the sum of employer and employee contributions across types of plans.

premiums for each type of coverage (single, employee-plus-one, or family coverage; see figure 6, panel B), as well as the share of employees who are policyholders in each type of coverage (Blewett and others 2019). We report average premiums across employees by weighting each coverage type by its employee share. Average health insurance premiums in 2019 were about $12,000. The average premium varied from about $7,000 for a single coverage plan, to $14,000 for plus-one coverage, to $20,000 for family coverage.[14]

On average, employers paid about three-quarters of these premiums (or $9,000 relative to the total average premium of $12,000). Gruber and McKnight (2003) discuss possible incentives for an employer to require some employee contributions to premiums, including encouraging those with outside insurance options to avail themselves of those instead or allowing firms to offer a range of low-cost plan options with employees contributing on the margin if they want more comprehensive coverage.[15]

II. Conceptual Framework

We sketch a simple, stylized model of a competitive labor market and use it to analyze qualitatively the equilibrium labor market impact of (1) counterfactually financing health insurance premiums through a proportional payroll tax rather than a fixed, per worker head tax and (2) counterfactually lower levels of health care spending. In the next sections, we calibrate a generalized version of this model for our quantitative analyses.

II.A. Setup

There are two types $g \in \{N, C\}$ of workers, where N and C denote noncollege and college workers, respectively. We assume that the productivity of college workers (A_C) exceeds that of noncollege workers ($A_C > A_N$). For the qualitative analysis, we assume a linear production technology for

14. The Medical Expenditure Panel Survey data used in figure 6 are not the only source of information on the cost to employers of providing health insurance. The Bureau of Labor Statistics provides an alternative estimate each quarter in the Employer Cost for Employee Compensation (ECEC), which includes the cost per employee hour of health benefits to the employer. In figure A.1 in the online appendix we adjust both estimates to make them comparable and compare their implications for the employer cost of providing health insurance. They line up quite closely.

15. Most employees are able to make their contributions out of pretax dollars. Gruber (2011) estimates that roughly 80 percent of employees with employer-provided coverage have access to a section 125 or cafeteria plan that allows them to make their health insurance contributions pretax.

output Y such that $Y = A_N L_N + A_C L_C$, where L_g is the employment of group g. The assumption that college and noncollege workers are perfect substitutes with a constant relative wage determined by their relative productivity irrespective of labor supply simplifies some of our comparative statics without affecting the qualitative insights that we emphasize. We also assume that everyone in the workforce holds employer-provided health insurance. We relax both assumptions in our quantitative calibration in the next sections.

LABOR DEMAND We consider a representative firm in a competitive labor market.[16] The firm chooses labor inputs to solve:

$$(1) \qquad max_{L_N, L_C} Y - \omega_N L_N - \omega_C L_C,$$

where w_g denotes the total cost to the employer of hiring a worker from group g. Under head tax financing of health insurance, $\omega_g = w_g + \tau$, where w_g is the salary paid to group g and τ is the health insurance premium per worker.[17] Under payroll tax financing, the cost of hiring a worker from group g is given by $\omega_g = (1 + t)\, w_g$, where t is the payroll tax on earnings. Given the assumption of a linear production technology, equilibrium costs per worker are determined by worker productivity. Thus, $\omega_g = A_g$.[18]

LABOR SUPPLY Each worker i in group $g \in \{N, C\}$ faces a discrete choice of whether or not to work. We model the indirect utility from employment (U^e_{gi}) as consisting of a systematic component V_g shared by all individuals in the group and an idiosyncratic component ε_i. The systematic component V_g depends on wages and health insurance provision. The idiosyncratic component $\varepsilon_i \geq 0$ captures the individual-specific disutility from working. This construction allows us to represent the indirect utility from working by $U^e_{gi} = V_g - \varepsilon_i = w_g + \alpha_g h - \varepsilon_i$, where $\alpha_g \geq 0$ is the group-specific amenity value of health insurance expenditures h relative to wages.

16. Focusing on a representative firm abstracts away from important ways in which firms influence labor market inequality (Card 2022).

17. Another important simplifying assumption is that there is an equal health insurance premium per worker at a representative firm. In reality, different firms may offer different benefit packages. In recent work, Ouimet and Tate (2023) and Lurie and Miller (2023) show that higher-income employees tend to select more-expensive plans. Indeed, they find that those in the top decile of the income distribution collectively account for 20 percent of total employer-sponsored health insurance premiums. That said, they find that inequality in premiums is much smaller than inequality in income.

18. The assumption that workers' compensation is equal to the value of their marginal product of labor is common but not innocuous. There are many reasons (e.g., imperfect competition and rent sharing) why worker compensation might not equal worker productivity.

A priori, α_g may be larger than or less than one. If health insurance is only available through the employer, employee risk aversion could produce a value of health insurance that is more than wages ($\alpha_g > 1$). In the presence of moral hazard, the privately and socially optimal amount of insurance would be to provide health insurance until $\alpha_g = 1$ (Baily 1978; Chetty 2006). However, as emphasized by Feldstein (1973), the preferential tax treatment of employer-provided health insurance can result in an equilibrium value of health insurance that is less than wages ($\alpha_g < 1$). When we keep constant the provision of employer-provided health insurance and focus solely on the impact of how it is financed, we can remain agnostic on the potentially group-specific utility to workers from health insurance relative to earnings, α_g.

We normalize the indirect utility from not working to zero. An individual will therefore work if and only if $U^e_{gi} > 0$. If we consider the idiosyncratic component of utility from work ε_i to be a random variable, the share of individuals in group g who will work is therefore given by:

$$P_g = Pr\left(V_g > \varepsilon_i\right) = Pr\left(w_g + \alpha_g h > \varepsilon_i\right).$$

For simplicity, we assume that the idiosyncratic components ε_i are independently and identically distributed according to a uniform distribution over the interval from $\underline{\kappa}$ to $\bar{\kappa}$ (i.e., $\varepsilon_i \sim U[\underline{\kappa}, \bar{\kappa}]$). The resulting aggregate labor supply function is therefore linear between $\underline{\kappa}$ and $\bar{\kappa}$:

$$P_g = Pr\left[V_g > \varepsilon_i\right] = \begin{cases} \dfrac{V_g - \underline{\kappa}}{\bar{\kappa} - \underline{\kappa}} & \text{if } V_g \in \left[\underline{\kappa}, \bar{\kappa}\right] \\ 0 & \text{if } V_g < \underline{\kappa} \\ 1 & \text{if } V_g > \bar{\kappa} \end{cases}. \tag{2}$$

We denote by $L_g = P_g N_g$ the total employees in group g, where N_g is the number of workers in group g.

II.B. Comparative Statistics: Head Tax vs. Payroll Tax Financing

Under the status quo head tax financing, the cost of hiring a worker is given by $\omega_g = w^H_g + \tau$, where we use the superscript H to denote the head tax scenario. Equilibrium wages are therefore $w^H_g = A_g - \tau$, and the college wage premium is:

$$\frac{w_C^H}{w_N^H} - 1 = \frac{A_C - \tau}{A_N - \tau} - 1. \tag{3}$$

The greater the health insurance premiums (τ) and the greater the relative productivity of college workers $\left(\frac{A_C}{A_N}\right)$, the greater the college wage premium. Substituting equilibrium wages into the labor supply function gives us relative equilibrium employment under the head tax (where $h = \tau$):

$$\frac{L_C^H}{L_N^H} = \frac{A_C + \left(\alpha_C - 1\right)\tau - \underline{\kappa}}{A_N + \left(\alpha_N - 1\right)\tau - \underline{\kappa}}. \tag{4}$$

We now consider the impact of financing employer-provided health insurance with a proportional, nationwide payroll tax instead of the current head tax. In this comparison we hold constant at τ the average employer contribution to each employee's health insurance. Thus, the per employee cost of health insurance remains unchanged at the national level; the only thing that changes is how that contribution is financed. As noted at the beginning of the paper, we also hold fixed the share of full-time, full-year workers receiving health insurance in this manner. This assumption allows us to focus exclusively on the impact of changing how a fixed amount of coverage is financed. This counterfactual also allows us to abstract from difficult-to-measure parameters such as the amenity value of health insurance premiums relative to wages (α_g), which affects equilibrium employment and (once we relax the assumption of a linear production technology) equilibrium wages as well.[19] More generally, the amenity value of employer-provided benefits is an important parameter for a range of policy counterfactuals (Gruber 1994; Summers 1989).

Because we assume that the health insurance provided remains constant under the counterfactual payroll tax financing, the labor supply function, equation (2), is unaffected. Labor demand, however, is affected. Under payroll tax financing (which is levied statutorily on firms), the employer now contributes a fixed proportion t of each worker's earnings to their health insurance, so that the cost per worker in group g is now $\omega = (1 + t)w_g^P$ where we use the superscript P to denote the payroll tax scenario. Equilibrium wages are now $w_g^P = \frac{A_g}{1 + t}$.

19. See equation (4) and online appendix A.2.

Since, by construction, the average per employee cost of health insurance is held constant, the payroll tax t is determined in equilibrium by the following budget constraint:

$$\tau \cdot \left(L_N + L_C\right) = t \cdot \left[w_N \cdot L_N + w_C \cdot L_C\right].$$

Given equilibrium wages under the payroll tax we can rewrite this relationship as:[20]

$$\tau = \left(\frac{t}{1+t}\right)\tilde{A}, \tag{5}$$

where $\tilde{A} \equiv \left(A_N \cdot \frac{L_N}{L_N + L_C} + A_C \cdot \frac{L_C}{L_N + L_C}\right)$ is the average productivity of college and noncollege workers weighted by their relative shares. Since $A_N < \tilde{A} < A_C$, it follows that the equilibrium wage for college workers is lower under payroll tax financing than head tax financing:

$$w_C^P = \frac{A_C}{1+t} < A_C - \tau = w_C^H. \tag{6}$$

By the same token, the equilibrium wage for noncollege workers is higher under payroll tax financing than under head tax financing:

$$w_N^P = \frac{A_N}{1+t} > A_N - \tau = w_N^H. \tag{7}$$

Because labor supply is the same under these two alternative financings, equilibrium employment under payroll tax financing is higher for noncollege workers ($L_N^P > L_N^H$) and lower for college workers ($L_C^P < L_C^H$).

Thus, the switch from head tax financing to payroll tax financing unambiguously reduces labor market inequality. The size of the health insurance premium τ compared to the gap in productivity between groups is key for determining the quantitative impact of the change in financing on labor market inequality. Note, however, that the effect on total employment

20. We assume that for determining labor demand, each firm takes the payroll tax t as fixed and ignores the impact of its hiring of noncollege and college workers on the equilibrium payroll tax. This seems reasonable given that any given firm's hiring has a negligible effect on the nationwide employment rates for these two types of workers.

is ambiguous since employment is increasing for noncollege workers and decreasing for college workers.

Figure 7 illustrates the impact of moving from head tax to payroll tax financing graphically. We plot the supply and demand for noncollege and college labor services separately in panels A and B, respectively. In panel C, we plot supply and demand of college relative to noncollege labor services in log terms to show directly how the college premium changes.[21] Because relative labor supply is unaffected by the form of financing, the shift from head tax to payroll tax financing lowers the college wage premium and lowers the college wage share, thereby reducing labor market inequality.

II.C. Comparative Statics: Counterfactual Levels of Health Care Spending

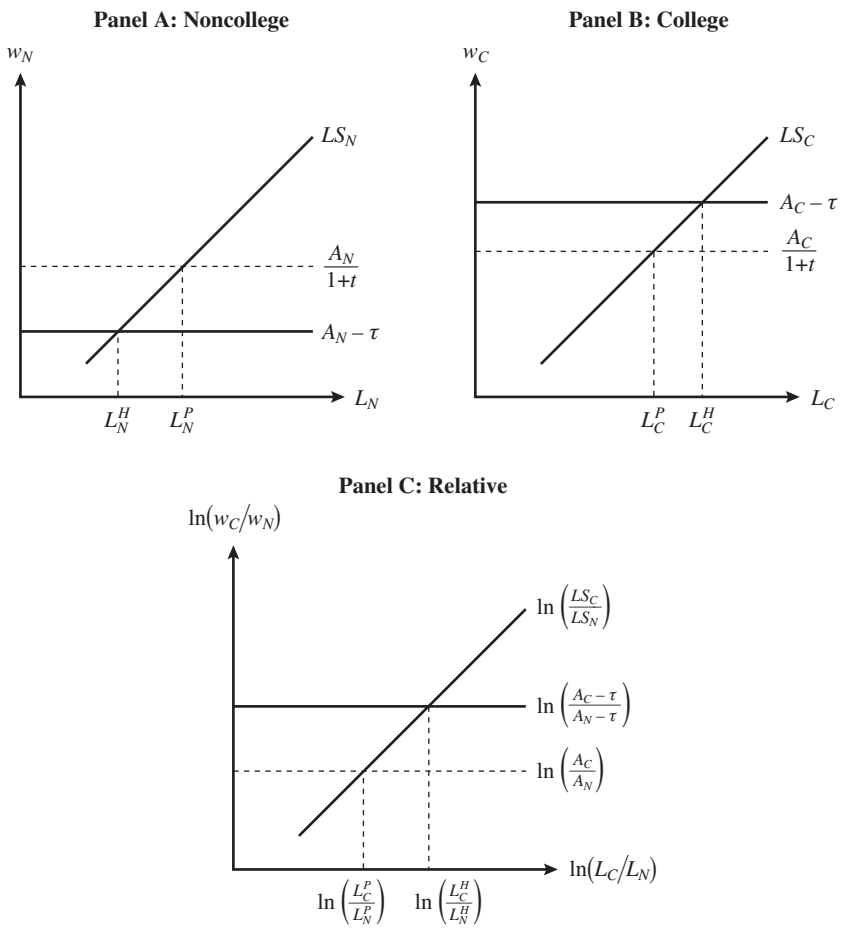
We can also use this framework to consider labor market outcomes under the current head tax regime but with counterfactual levels of health care spending and hence counterfactual levels for the head tax τ. A decline in τ shifts out labor demand (since the total cost for hiring an employee is smaller) and also shifts in labor supply (since less health coverage is provided at a given wage).[22] As can be seen from equation (3), these shifts would reduce the college wage premium. To determine how equilibrium employment outcomes would change, however, we must make additional assumptions regarding the workers' amenity value α_g of employer-provided health insurance relative to cash. These assumptions were not required for our analysis of counterfactual financing of employer-provided health insurance since the amount of health insurance provided was held constant under the counterfactual.[23] They are required for this counterfactual since

21. To obtain the log linear relative labor supply expression depicted in panel C, we log linearize our expressions for relative labor supply $\left(\frac{L_C}{L_N} = \frac{w_C + \alpha_C \tau - \underline{\kappa}}{w_N + \alpha_N \tau - \underline{\kappa}} \cdot \frac{N_C}{N_N}\right)$.

22. This counterfactual lowering of t can also be loosely viewed as a way of analyzing the impact of the Cadillac tax that was enacted but never implemented, as discussed at the beginning. In particular, one way of interpreting a goal of the Cadillac tax is that it was designed to reduce the generosity of health insurance provision. However, the effects on disposable income and labor supply of different groups, whose marginal tax rates vary, would need to be incorporated into our model to characterize the effects of a Cadillac tax more realistically. In addition, there are several other important aspects of this policy that are beyond the scope of our paper, such as the prevalence of Cadillac health plans across different groups.

23. Specifically, as discussed in online appendix A.2, the amenity value α_g is relevant only for pinning down the intercept of the labor supply function. In our payroll tax analysis, we assume this amenity value is the same across groups ($\alpha_C = \alpha_N$); in our sensitivity analysis, where we make a direct assumption about the shape of the labor supply function, we also do not have to make any additional assumptions about α_g.

Figure 7. Moving from a Head Tax to a Payroll Tax Equilibrium



Source: Authors' calculations.

Note: This figure shows the impact on equilibrium wages and employment from switching from a head tax (τ) to a payroll tax (t). Panel A shows the results for noncollege educated workers (with productivity A_N), panel B shows the results for college-educated workers (with productivity A_C), and panel C shows relative outcomes for college workers relative to noncollege workers. We use superscript H to denote outcomes under the head tax and superscript P to denote outcomes under the payroll tax. As shown in panel C, the change from head tax to payroll tax financing unambiguously lowers the college wage premium and college employment relative to noncollege employment.

they determine the magnitude of labor supply shifts for both groups. If the amenity value of rising health care spending were zero, rising health care spending would function in the same way as a standard tax increase, with impacts on wages and employment that depend on relative supply and demand elasticities. If the amenity value were one for both groups, labor supply would shift out by the same amount as the cost increase and effects would be entirely on wages, with no impact on employment (Summers 1989; Gruber 1994).

Figure 8 illustrates the impact of a reduction in the cost of employer-provided health insurance from τ to τ' when assuming the amenity value of this insurance is equal to one for both groups (i.e., assuming $\alpha_N = \alpha_C = 1$). Once again, panels A and B plot the demand for and supply of labor for noncollege and college workers, respectively, while panel C shows the relative supply and demand of college versus noncollege labor services in log terms. For both groups of workers, the reduction in the cost of employer-provided health insurance leads to an increase in labor demand at any given wage and so wages unambiguously increase for both groups. Since wages increase by a similar amount for both groups, wage inequality decreases as shown in panel C. However, the effect on labor supply depends on the amenity value of employer-provided health insurance. Since we assumed $\alpha = 1$ for both groups, the shift in labor demand is matched exactly by a shift in labor supply so that the quantity of labor supplied is unaffected. For $\alpha < 1$, the shift in labor supply would be less than the shift in labor demand and labor supply would increase for both groups; conversely for $\alpha > 1$, the shift in labor supply would be greater than the shift in labor demand and labor supply would decrease for both groups.

III. Calibration and Implied Parameter Values

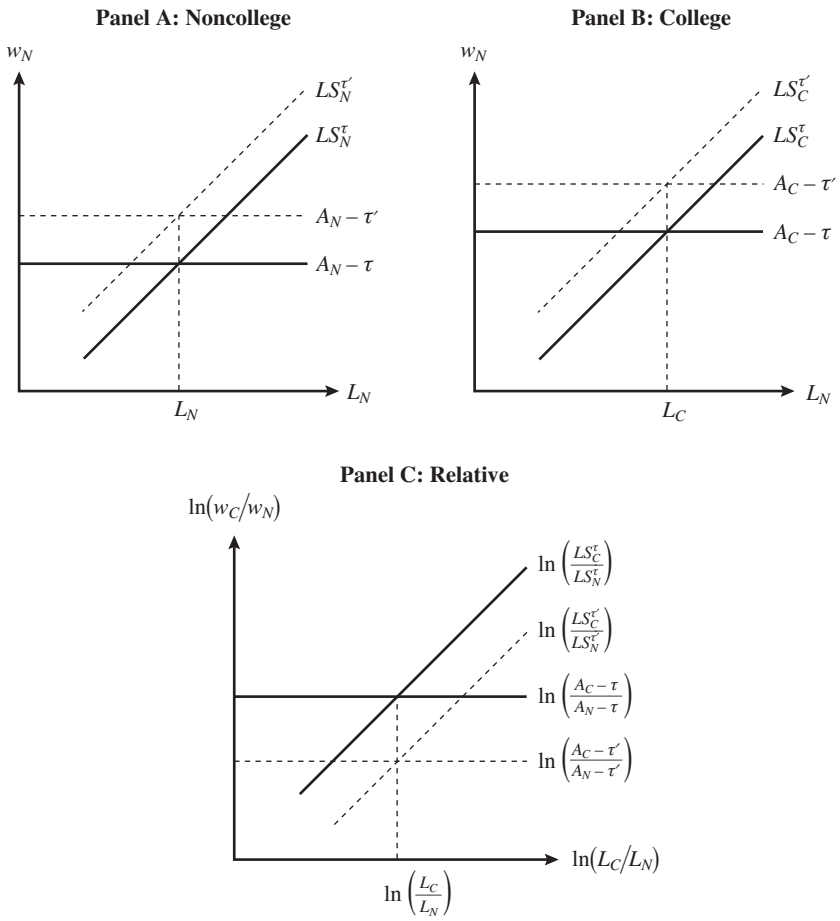
III.A. Calibration

For our calibration exercise, we generalize the linear technology used in our qualitative analysis to a CES production function:

$$Y = \left(\lambda_N L_N^{\rho} + \lambda_C L_C^{\rho}\right)^{1/\rho},$$

where λ_g is a group-specific productivity shifter, and the parameter $\rho < 1$ dictates the relative substitutability or complementarity of noncollege and college workers. When $\rho = 1$, this gives us the linear production function discussed before and implies that the two types of workers are perfect

Figure 8. Reducing the Cost of Health Insurance



Source: Authors' calculations.

Note: This figure shows the impact on equilibrium wages and employment from reducing τ to τ' under the assumption of $a_g = 1$. Panel A shows the results for noncollege-educated workers (with productivity A_N), panel B shows the results for college-educated workers (with productivity A_C), and panel C shows relative outcomes for college workers relative to noncollege workers. As shown in panel C, the reduction in τ unambiguously lowers the college wage premium.

substitutes. For our baseline analysis, we assume $\rho = .038$ based on Autor, Goldin, and Katz (2020), and explore sensitivity to other assumptions below.

Given per worker costs ω_g, the firm chooses group-specific labor inputs to maximize profits:

$$max_{L_N, L_C}\left(\lambda_N L_N^{\rho} + \lambda_C L_C^{\rho}\right)^{1/\rho} - \omega_N L_N - \omega_C L_C. \tag{8}$$

We assume that under the head tax regime, equilibrium wages and employment for each group are given by their observed 2019 values (table 1).

A key calibration choice concerns the value of the head tax τ. Average premiums for employer-provided health insurance were \$11,764 in 2019 (figure 6). However, as seen in table 1, not all full-time, full-year workers are enrolled in employer-provided health insurance. In online appendix A.3, we show that in a model of incomplete take-up, the effective τ is simply the average premium scaled by the share of employees who are policyholders.[24] Since only 66 percent of full-time, full-year workers are policyholders, for our baseline analysis we therefore scale down average premiums by 0.66. This gives us a baseline value of $\tau = \$7{,}758$ in 2019.

III.B. Implied Parameter Values

With values for L_g, w_g, ρ, and τ, this setup allows us to solve for the remaining model parameters: the productivity shifters λ_C and λ_N and the value of $(\bar{\kappa} - \underline{\kappa})$, which governs the slope of the labor supply function. Specifically, we can solve the firm's maximization problem in equation (8) for the productivity shifters λ_C and λ_N, and we use the labor supply function in equation (2) to solve for $(\bar{\kappa} - \underline{\kappa})$. Intuitively, the productivity shifters λ_C and λ_N are revealed by the firm's first-order conditions for labor demand, and the labor supply slope $(\bar{\kappa} - \underline{\kappa})$ is identified from our assumption that the observed equilibrium wage and employment allocations for each group are produced by a linear labor supply function with a common slope. Online appendix A.2 provides the derivation.

Having identified the baseline parameters of the CES production function and the slope of the labor supply function, we can then solve for

24. In practice, table 1 showed that the lack of coverage reflects—in roughly equal measure—the fact that some firms do not offer health insurance and that some workers who are offered do not take up the insurance. Modeling incomplete offering of health insurance is more complicated, as we must then solve for equilibrium sorting of workers across firms that do and do not offer insurance. We discuss some of the implications of this possible extension in the last section.

the equilibrium under the payroll tax, where the cost per worker is now $\omega_g^P = (1 + t) \cdot w_g^P$. To do so, we use the solution to the firm's maximization problem in equation (8), together with the labor supply function in equation (2), and the equilibrium condition for the payroll tax.[25]

The equilibrium condition for the payroll tax that generalizes equation (5) for the CES production function is given by:

$$t = \frac{\tau}{\tilde{w} - \tau},$$

where $\tilde{w}$ is the average wage, which equals $\frac{L_N}{L_N + L_C} \cdot w_N^P + \frac{L_C}{L_N + L_C} \cdot w_C^P$. Together, this equilibrium condition gives us five equations—first-order conditions for each group's employment, labor supply functions for each group, and the equilibrium payroll tax—for the five unknowns. This system of equations allows us to solve for wages and employment of each group, as well as the payroll tax using a nonlinear equation solver.

Table 2 provides the baseline implied values for key model parameters under the baseline assumed value of τ = \$7,758 and the assumed substitution parameter $\rho = 0.38$. The college productivity shifter λ_C is roughly 50 percent higher than that of noncollege workers λ_N.[26] Our estimate of $(\bar{\kappa} - \underline{\kappa}) = 316{,}743$ can be translated into an implied labor supply elasticity

25. Specifically, since we know the slope of the labor supply function, we use a modified version of this equation that uses the head tax equilibrium as an intercept and calculates the labor supply given the slope of labor supply and the change in wages. Full details provided in online appendix A.2.

26. To provide some intuition for these productivity shifters, recall that when $\rho = 1$, the firm problem simplifies to equation (1) and then $\lambda_g = A_g$. Given the value of college and noncollege earnings in table 1, together with the assumed value of τ = 7,883, this implies that A_C = 104,062 and A_N = 57,937. Thus, in the case of perfect substitutes, we get that $\frac{\lambda_C}{\lambda_N} = 1.80$. Our baseline calibration of $\rho = 0.38$ has a production technology with less substitutability. The effect of this substitutability on the value of λ_C and λ_N can be seen from investigating the first-order conditions for labor. The ratio of first-order conditions for noncollege and college workers gives the following expression: $\lambda_N = \left(\frac{w_N + \tau}{w_C + \tau}\right) \cdot \left(\frac{L_N}{L_C}\right)^{1-\rho} \cdot \lambda_C$. Rearranging and taking logs gives an expression $ln\left(\frac{w_C + \tau}{w_C + \tau}\right) = ln\frac{\lambda_C}{\lambda_N} - (1 - \rho) ln\frac{L_C}{L_N}$ for the log college wage premium, which is determined by a race between education and technology (Goldin and Katz 2008). The $ln\frac{\lambda_C}{\lambda_N}$ term is the intercept or technology term of the relative inverse demand for college workers.

Table 2. Baseline Implied Parameter Values

College productivity shifter	λ_C	59.584
Noncollege productivity shifter	λ_N	38.794
Difference in reservation wages	$(\bar{\kappa} - \underline{\kappa})$	\$316,743

Source: Authors' calculations.

Note: Our baseline analysis assumes, following Autor, Goldin, and Katz (2020), a substitution parameter (ρ) of 0.38 and that the cost of employer-provided health insurance (τ) is \$7,758. Dollar amounts are in 2019 US dollars.

of $\frac{1}{\left(\bar{\kappa} - \underline{\kappa}\right)} \cdot \frac{w_g}{P_g} = \frac{1}{316,743} \cdot \frac{w_g}{P_G}$ for group g. Given the observed equilibrium values of w_g and P_g in table 1, the implied elasticities are 0.40 and 0.26 for college and noncollege groups, respectively. These estimates are within the range of estimates in the literature. For example, Chetty (2012) reports that estimates of extensive margin elasticities range from 0.15 to 0.45. Chetty and others (2013) provide a meta-analysis that points to an extensive margin labor supply elasticity of around 0.25, though estimates based on macroeconomic data tend to be larger (Keane and Rogerson 2012; Mui and Schoefer 2021).

IV. Labor Market Outcomes under Counterfactual Payroll Tax Financing

IV.A. Effects of Payroll Tax Financing on 2019 Labor Market Outcomes

Table 3 shows how labor market outcomes would differ in 2019 if, counterfactually, employer-provided health insurance were financed by a proportional payroll tax rather than a fixed, per worker head tax. The first column shows our baseline analysis, which assumes τ = \$7,758. This approach scales down the average \$11,764 health insurance premium to account for the fact that only about two-thirds of workers are policyholders.[27] With this τ, we calculate that the counterfactual, equilibrium payroll tax t would be 11 percent. That is, switching to payroll tax financing would add an additional 11 percentage points to existing payroll and income tax rates.

27. Because the cost of hiring a worker depends on the total health insurance premium, regardless of its statutory incidence, our calculation of τ does not adjust for the fact that only about three-quarters of premiums are paid by the employer (figure 6). This calculation follows the standard approach in public finance to disregard statutory incidence in calculating the economic incidence of taxes; in our setting, as discussed above, the split into employer and employee contributions is likely an equilibrium response.

Table 3. 2019 Labor Market Effects of Counterfactual Payroll Tax Financing

	(1) Baseline	*(2) Full coverage*
Fixed per worker cost, τ	$7,758	$11,764
Payroll tax rate, t	11.06%	16.80%
Wages		
Change in college wage, $\Delta(w_C)$	–$2,181	–$3,158
Change in noncollege wage, $\Delta(w_N)$	$1,660	$2,383
Change in college wage premium (%)	–11.26	–16.00
Employment		
Change in college employment rate, $\Delta(P_C)$	–0.69 pp	–1.00 pp
Change in noncollege employment rate, $\Delta(P_N)$	0.52 pp	0.75 pp
Change in total employment, $\Delta(L)$	86,833	119,495
Change in college employment, $\Delta(L_C)$	–408,588	–591,747 pp
Change in noncollege employment, $\Delta(L_N)$	495,420	711,242
Change in college share of wage bill, $\Delta\left(\frac{w_C L_C}{w_N L_N + w_C L_C}\right)$	–1.77 pp	–2.55 pp

Source: Authors' calculations.

Note: This table presents the change in 2019 outcomes from moving to counterfactual payroll tax financing relative to the current head tax financing. Each column shows results under a different measure of the cost of employer-sponsored health insurance (τ). The first column shows results of our baseline estimate: total premium accounting for the fact that only 66 percent of workers are policyholders. The second column shows results for the total premium cost with the assumption that all workers eligible for employer-provided health insurance are policyholders. The college wage premium (CWP) is defined by $CWP = w_C/w_N - 1$ and often referred to as a percentage, that is, a value of 0.90 implies college workers make 90 percent more than noncollege workers. In this table, the percentage change in the college wage premium is calculated as the percentage change in this value when moving from the head tax to the payroll tax, which would be equal to $(CWP^P - CWP^H)/CWP^H$. All dollars are in 2019 US dollars.

We estimate that if employer-provided health insurance were financed by this payroll tax, the wages of college graduates would fall by $2,181, the wages of noncollege graduates would rise by $1,660, and the college wage premium would fall by 11.1 percent. Employment would increase by 86,833 jobs, with an increase of 495,420 jobs for noncollege workers that is offset by 408,588 fewer college jobs. These wage and employment changes would result in a 1.8 percentage point lower college share of the wage bill.

For comparison, the second column shows results under the assumption that the head tax τ is equal to the average $11,764 health insurance premium for employer-provided health insurance (which naturally changes the implied parameter values in table 2). This value for τ corresponds in a sense to the raw policy instrument: it is the cost of providing all full-time, full-year workers with employer-provided health insurance financed through a head tax. The counterfactual analysis thus tells us the impact of instead providing all of these workers, rather than only the 66 percent who are policyholders, with the same health insurance financed through a payroll

tax. In this case, we estimate that the counterfactual equilibrium payroll tax rate would be 16.8 percent, rather than 11 percent in our baseline, and the college wage premium would fall by 16 percent, rather than 11.1 percent. Likewise, our baseline estimate of the approximately 500,000 increase in noncollege employment would increase to 710,000 under this alternative assumption.

SENSITIVITY ANALYSIS Table 4 holds the value of t constant at our baseline of $7,758 and shows the sensitivity of our results to other assumptions. As shown in the last column, these alternative assumptions have virtually no impact on the equilibrium payroll tax rate, which ranges from 11.05 percent to 11.07 percent. They cause only slight adjustments to the other equilibrium outcomes.

Panel A shows results under different labor supply elasticities. Recall that for our baseline we derived group-specific elasticities based on the observed equilibrium allocations for each group and the assumption of a common labor supply slope across groups (i.e., κs are not group-specific). Here, we provide results based on directly calibrating a common labor supply elasticity. Since Chetty (2012) reports estimates of extensive margin elasticities ranging from 0.15 to 0.45, we show results for 0.15, 0.3, and 0.45. When we assume a common labor supply elasticity across groups, we now allow the κ parameters to be group-specific. In other words, to implement counterfactuals that calibrate labor supply elasticities directly, we set the slope of the labor supply functions for both groups to be consistent with the desired labor supply elasticities.[28] When labor supply elasticities are lower, the wage premium effects are bigger in absolute value, and changes in the college wage bill share and noncollege employment are smaller in absolute value. Intuitively, more of the equilibrium adjustment to the financing change happens via prices rather than quantities when labor supply is less elastic. Depending on this assumption, the decrease in the college wage premium ranges from 10.5 to 12.2 percent.

Panel B shows results using higher and lower assumptions about substitutability across worker types relative to our baseline level of substitutability $\rho = 0.38$ from Autor, Goldin, and Katz (2020). Alternative assumptions about substitutability change the slope of the relative labor demand curve

28. Specifically, given the assumed elasticity ε_g, we use the expression for the elasticity $\varepsilon_g = \frac{1}{\left(\overline{\kappa_g} - \underline{\kappa_g}\right)} \cdot \frac{w_g}{P_g}$ to identify the group-specific slope of the labor supply function $\left(\overline{\kappa_g} - \underline{\kappa_g}\right) = \frac{1}{\varepsilon_g} \cdot \frac{w_g}{P_g}$. Other than this detail, calibration is the same as in the baseline analysis.

Table 4. Sensitivity Analysis: 2019 Labor Market Effects of Counterfactual Payroll Tax Financing

	Change in college wage premium (%)	*Change in noncollege wages* $\Delta(w_N)$	*Change in college employment rate* $\Delta(P_C)$	*Change in noncollege employment rate* $\Delta(P_C)$	*Change in noncollege employment (thousands)* $\Delta(L_N)$	*Payroll tax rate t (%)*
Baseline	−11.26	$1,660	−0.69 pp	0.52 pp	495.42	11.06
Panel A: Labor supply elasticities						
Derived group-specific elasticities						
$\varepsilon_C = 0.42$ and $\varepsilon_N = 0.28$ (baseline)	−11.26	$1,660	−0.69 pp	0.52 pp	495.42	11.06
Assumed common elasticities						
$\varepsilon_C = \varepsilon_N = 0.15$	−12.28	$1,824	−0.28 pp	0.34 pp	317.53	11.05
$\varepsilon_C = \varepsilon_N = 0.30$	−11.35	$1,674	−0.52 pp	0.62 pp	582.94	11.06
$\varepsilon_C = \varepsilon_N = 0.45$	−10.55	$1,547	−0.73 pp	0.85 pp	807.90	11.07
Panel B: Substitutability (ρ)						
Perfect substitutes ($\rho = 1$)	−13.39	$1,985	−0.82 pp	0.63 pp	592.36	11.07
Gross substitutes ($\rho = 0.38$, baseline)	−11.26	$1,660	−0.69 pp	0.52 pp	495.42	11.06
Cobb-Douglas ($\rho = 0.01$)	−10.28	$1,511	−0.63 pp	0.48 pp	450.92	11.06

Source: Authors' calculations.

Note: This table shows the sensitivity of our baseline analysis (table 3, column 1; τ = $7,758) of the impact of moving to counterfactual payroll tax financing on 2019 outcomes. The college wage premium is defined as $CWP = w_C/w_N - 1$. In this table, the percentage change in the college wage premium is calculated as the percentage change in this value when moving from the head tax to the payroll tax, which would be equal to $(CWP^P - CWP^H)/CWP^H$.

for college workers compared to noncollege workers. Intuitively, as that substitutability increases, the impact of a given health financing tax on employment also increases. However, in practice, because our assumed value of ρ affects the calibration of the productivity shifters, our exercise is not one of pure comparative statics. Nonetheless, we find that this qualitative intuition holds. Thus, if college and noncollege workers were perfect substitutes, as in section II, with an assumed linear production function (i.e., $\rho = 1$), the changes in outcomes would be larger in absolute value: the college premium would decline by 13 percent, and the noncollege employment rate would increase by 592,000 instead of 495,000. If the production technology combining college and noncollege workers were nearly Cobb-Douglas (i.e., $\rho = 0.01$), changes in outcomes would be smaller in absolute value, with a 10 percent increase in the college wage premium and an increase in noncollege employment of 451,000.

In online appendix B we also show results when we expand our definition of college-educated workers to include those with some college education, even if they do not have a bachelor's degree.

IV.B. Effects of Payroll Tax Financing on Changes in Labor Market Outcomes, 1977–2019

Table 5 considers what changes in labor market inequality from 1977 to 2019 would have looked like if payroll tax financing had been in place throughout. We use the observed values of τ (figure 6) in 1977 and 2019 in the baseline calculations and assume throughout that 66 percent of full-time, full-year workers are policyholders of employer-provided health insurance.[29] The first column shows changes over time that have occurred under the head tax regime. Column 2 shows counterfactual changes under our baseline assumption of τ, which scales average premiums by 0.66 to reflect the share of workers that are policyholders. Column 3 shows results without that scaling. We focus our discussion on the baseline results.

The results show that labor market outcomes would have evolved quite differently under payroll tax financing. Under the head tax regime, real college wages (in 2019 dollars) increased by $33,903, while real noncollege wages grew by $7,754. As a result, the college premium increased by 44.8 percentage points over this period. Employment rates also increased for both groups; the employment rate increased by 5.8 percentage points for college workers and by 9.1 percentage points for noncollege workers.

29. In practice, this number ranges from 0.73 to 0.66 from 1996 to 2019, but since it is not available for every year, we use the statistic for 2019 throughout.

Table 5. Changes over Time: Labor Market Effects of Counterfactual Payroll Tax Financing, 1977–2019

	Payroll tax equilibrium		
	(1) Head tax equilibrium	*(2) Baseline*	*(3) Full coverage*
Employer-sponsored health insurance			
Change in cost ($\tau_{2019} - \tau_{1977}$)	—	$5,937	$9,003
Payroll tax ($t_{2019} - t_{1977}$)	—	7.16 pp	10.88 pp
Wages			
Change in college wages ($w_{C,2019} - w_{C,1977}$)	$33,903	$32,121	$31,339
Change in noncollege wages ($w_{N,2019} - w_{N,1977}$)	$7,754	$9,305	$9,976
Change in college wage premium (percentage points)	44.83	35.80	32.08
Employment rate			
Change in college employment rate ($P_{C,2019} - P_{C,1977}$)	5.77 pp	5.44 pp	5.31 pp
Change in noncollege employment rate ($P_{N,2019} - P_{N,1977}$)	9.13 pp	9.55 pp	9.73 pp
Change in college share of wage bill $\left(\frac{w_C L_C}{w_N L_N + w_C L_C}\right)_{2019} - \left(\frac{w_C L_C}{w_N L_N + w_C L_C}\right)_{1977}$	31.06 pp	29.62 pp	29.00 pp

Source: Authors' calculations.

Note: This table presents the change of each outcome between 1977 and 2019 for the head tax equilibrium and for the payroll tax counterfactual. Columns 2 and 3 show results for different measures of the cost of employer-sponsored health insurance (τ). The college wage premium is defined as $CWP = w_C/w_N - 1$ and often referred to as a percentage, that is, a value of 0.90 implies college workers make 90 percent more than noncollege workers. In this table, the percentage point change in the college wage premium is calculated as the percentage point change in this value from 1977 to 2019, which would be equal to $CWP_{2019} - CWP_{1977}$.

The payroll tax rate necessary to finance health insurance at the same per capita cost as under the head tax would have increased steadily from around 4 percent in 1977 to 11 percent in 2019. Under this alternative financing, the increase in the college wage premium would have been about 20 percent smaller: a 35.8 percentage point increase instead of the observed 44.8 percentage point increase. The employment rate of noncollege workers would have increased by about 4.5 percent more; specifically, the noncollege employment rate would have increased by 9.6 percentage points under the payroll tax, instead of the observed 9.1 percentage points.

SENSITIVITY ANALYSIS Table 6 once again shows the sensitivity of these results to alternative assumptions about the substitutability of workers and labor supply elasticities. Our baseline estimate (table 5, column 2) that the college wage premium would have risen about 20 percent less varies

Table 6. Sensitivity Analysis: Changes over Time: Labor Market Effects of Counterfactual Payroll Tax Financing, 1977–2019

	Change in college wage premium	*Change in noncollege wages* ($w_{N,2019} - w_{N,1977}$)	*Change in college employment rate* ($P_{C,2019} - P_{C,1977}$)	*Change in noncollege employment rate* ($P_{N,2019} - P_{N,1977}$)	*Change in payroll tax rate* ($t_{2019} - t_{1977}$)
Head tax equilibrium	44.83 pp	$7,754	5.77 pp	9.13 pp	—
Payroll tax equilibrium					
Labor supply elasticities					
Derived group-specific elasticities					
$\varepsilon_C = 0.42$ and $\varepsilon_N = 0.28$ (baseline)	35.80 pp	$9,305	5.44 pp	9.55 pp	7.16 pp
Assumed common elasticities					
$\varepsilon_C = \varepsilon_N = 0.15$	35.32 pp	$9,430	5.58 pp	9.43 pp	7.15 pp
$\varepsilon_C = \varepsilon_N = 0.30$	36.04 pp	$9,292	5.42 pp	9.69 pp	7.16 pp
$\varepsilon_C = \varepsilon_N = 0.45$	36.66 pp	$9,175	5.28 pp	9.91 pp	7.17 pp
Substitutability (ρ)					
Perfect substitutes ($\rho = 1$)	34.46 pp	$9,578	5.47 pp	9.61 pp	7.17 pp
Gross substitutes ($\rho = 0.38$, baseline)	35.80 pp	$9,305	5.44 pp	9.55 pp	7.16 pp
Cobb-Douglas ($\rho = 0.01$)	36.49 pp	$9,174	5.44 pp	9.52 pp	7.16 pp

Source: Authors' calculations.

Note: This table shows the sensitivity of our baseline analysis (table 5, column 2) of the impact of moving to counterfactual payroll tax financing on the change in outcomes between 1977 and 2019. The college wage premium is defined as $CWP = w_C/w_N - 1$ and often referred to as a percentage; therefore, in this table, the percentage point change in the college wage premium is calculated as the percentage point change in this value from 1977 to 2019, which would be equal to $CWP_{2019} - CWP_{1977}$.

modestly from 18 to 23 percent less. Online appendix B shows results when workers with some college education are included in the definition of college-educated workers.

DISAGGREGATING EFFECTS BY SEX The employment rate for noncollege males declined by 4.3 percentage points from 1977 to 2019 (online appendix table A.5). To consider how this group would have fared under a payroll tax, we extend the baseline model to account for different trends in employment and wages by sex. To do so, we assume that male and female workers of the same group are perfect substitutes. Total labor input from group $g \in \{N, C\}$ is therefore equal to $L_g = X_{g,M} L_{g,M} + X_{g,F} L_{g,F}$, where we normalize the male-specific productivity shifter $X_{g,M}$ to one. We also assume that the parameters determining labor supply $\overline{\kappa_s}$ and $\underline{\kappa_s}$ are sex-specific. By disaggregating wage and employment rate data by sex, the rest of the calibration follows the same logic as our main analysis.

Online appendix table A.4 shows the effects of moving from a head tax to a payroll tax in 2019 in aggregate and for each sex. In aggregate, the results largely match those from our main analysis, yet differences emerge when disaggregated by sex. The payroll tax has a redistributive effect such that the wage loss is biggest for the subgroup with the highest wages, college males, and the wage gain is biggest for the subgroup with the lowest wages, noncollege females. Meanwhile, the effects for college females and noncollege males are relatively smaller. Changes in employment rates and employment follow a similar pattern.

Online appendix tables A.5 and A.6 show the changes over time under a head tax versus a payroll tax for males and females, respectively. As before, the payroll tax is redistributive such that the wage loss is biggest for college males and the wage gain is biggest for noncollege females, while college females and noncollege males experience more modest effects. Since the effect is more modest for noncollege males, their wage increase would only be $1,300 greater and their employment rate decline would be only 0.4 percentage points smaller from 1977 to 2019 under the payroll tax than the head tax. Therefore, the results suggest the health wedge is unlikely to account for much of the excess decline in employment rates for noncollege males.

IV.C. Benchmarking the Results

To benchmark these calibration results, we compare the impacts of counterfactual payroll tax financing to existing estimates from the literature of the impact of other factors on US labor market inequality. Although there is vast literature to draw on, direct comparisons are often hampered by differences across papers in the outcomes analyzed, the measures of inequality

used, and the time periods studied. Still, we are able to provide some benchmarks for our estimated impacts on the college wage premium, noncollege employment, and noncollege wages. Where we are able to make reasonable comparisons, they suggest that the magnitude of the health wedge effect rivals other leading causes of labor market inequality, including outsourcing, union density, trade, the relative supply of college graduates, and automation.

COLLEGE WAGE PREMIUM Our baseline estimate is that switching from head tax to payroll tax financing of employer-provided health insurance would decrease the 2019 college wage premium by 11 percent. Our sensitivity analyses suggest a range for this estimate of 10 to 13 percent. This decline in the college wage premium is comparable to the effect of outsourcing that Goldschmidt and Schmieder (2017) estimate. Specifically, they find that domestic outsourcing in Germany causes wages at the outsourcing establishment to fall about 10 percent for non-outsourced noncollege-educated workers, while there is no effect on wages of non-outsourced college-educated workers; this suggests that outsourcing raises the college premium at the parent establishment by about 10 percent.

Another way to benchmark our estimate is to consider the equivalent shock to the relative supply of college workers needed to cause the college premium to decline by the same magnitude as the impact of switching to payroll tax financing. Autor, Goldin, and Katz (2020), who update Katz and Murphy (1992), estimate that a 10 percent increase in the relative supply of college equivalents reduces the college premium by about 6 percent.[30] From 1979 to 2017, the log relative supply of college equivalents fell 0.213 per decade or 2.13 log points per year. A decade of decline in the relative supply of college equivalents would therefore increase the log college wage premium by $0.131 \left(= \frac{0.213}{1.62} \right)$. Thus, our estimate that the college wage premium would be around 11 percent lower under payroll tax financing is roughly the same magnitude (albeit opposite in sign) as the estimated impact of a decade of decline in the relative supply of college workers.

We can also compare our estimate of the decline in the college wage premium from payroll tax financing to estimates of the impact of the minimum wage on the college wage premium. In particular, Vogel (2022) generalizes the canonical model in which the college wage premium is

30. Note that college equivalents are college plus half those with some college, and noncollege workers are those with twelve years or fewer of schooling and half of those with some college.

determined by the relative supply and demand of labor to incorporate labor market institutions—such as monopsony power, minimum wages, and unemployment—and uses this framework to estimate the elasticity of the college wage premium with respect to the real minimum wage. These results suggest that the real minimum wage would need to increase by around 30 percent to achieve an equivalent of the 11 percent decrease in the college wage premium that we estimate would occur from switching from head tax to payroll tax financing.[31]

Much of the literature has focused on the role of various factors in contributing to changes in labor market inequality over time. We estimate that under counterfactual payroll tax financing, the college wage premium would have increased by 20 percent less from 1977 through 2019. In our sensitivity analyses, alternative calibrations suggest a range of 18 to 23 percent for how much less the college wage premium would have increased. This estimate is roughly comparable to the estimated impact of rising trade and declining unionization (albeit over slightly different time periods). Binder and Bound (2019) provide a useful summary of the literature on the contribution of rising trade and declining unionization to the change in the college premium between 1980 and 2006. They cite Katz's (2008) comment on Krugman (2008), suggesting that "rising trade accounted for less than 20 percent of the increase in the college wage premium" over this period (Binder and Bound 2019, 167). This assessment preceded the findings of Autor, Dorn, and Hanson (2013), which have shed further light on the effect of trade on inequality. Binder and Bound (2019) also note that DiNardo, Fortin, and Lemieux (1996) find the decline in unionization led to around 20 percent of the rise in the college wage premium over the 1980s. Using a related measure of inequality, Farber and others (2021) find declines in union density explain about 12 to 21 percent of the increase in the Gini coefficient from 1968 to 2014.[32]

NONCOLLEGE WAGES Our baseline estimate is that switching from head tax to payroll tax financing would increase the annual earnings of noncollege

31. Vogel (2022) measures the college wage premium as the ratio of wages instead of the percentage difference, but this benchmark calculation accounts for this difference. Additionally, note that the results in Vogel (2022) are based on the average minimum wage across states instead of the federal minimum wage.

32. Farber and others (2021) find that a 10 percentage point decline in union density—which is roughly the observed decline since 1980—reduced the college premium by about 12–15 percent when using time series variation, but they found somewhat smaller effects when using micro-level and state-year panel designs. Their state-year panel estimates, for example, were about half the magnitude of the time series estimates, and the micro estimates were in between those of the two designs.

full-time, full-year workers by about $1,700 (about 3 percent). Two useful benchmarks are the estimate of an average 10 percent union wage premium over the last two decades (Farber and others 2021) and the estimate that domestic outsourcing of workers in Germany causes wages to fall about 10 percent for noncollege-educated non-outsourced workers (Goldschmidt and Schmieder 2017).

Other useful comparisons are to the impact of increased exposure to imports or to robots. Chetverikov, Larsen, and Palmer (2016) estimate that for every $1,000 increase in import exposure per worker—roughly the interquartile range in CZ-level import exposure growth from 2000 through 2007 in Autor, Dorn, and Hanson (2013)—weekly wages decrease by about 0.7 percent on average, with workers in the bottom quartile of the wage distribution experiencing declines about twice as large in percentage terms and those at the top experiencing smaller declines. Acemoglu and Restrepo (2020) estimate that one more robot per thousand workers (the observed increase from 1993 to 2007) decreased wages by 0.4 percent, with effects again concentrated on those toward the lower end of the wage distribution. Our estimate of the impact on noncollege wages of switching to payroll tax financing is thus larger than the estimated impact of a $1,000 increase in import exposure per worker or of the increase in robots per thousand workers from 1992 to 2007.

NONCOLLEGE EMPLOYMENT Our baseline estimate is that switching from head tax to payroll tax financing of employer-provided health insurance would increase the 2019 employment rate for noncollege workers by 0.52 percentage points. In our sensitivity analyses, alternative calibrations suggest a range for the increase in the noncollege employment rate from 0.3 to 0.8. This effect is roughly comparable in magnitude to what others have estimated would be the decline in noncollege employment from a $500 increase in import exposure per worker, or a doubling of the growth in the number of robots per one thousand workers between 1993 and 2007.

To gauge the impact of the spread of robots, we look to Acemoglu and Restrepo (2020). They estimate that an additional robot per thousand workers between 1993 and 2007 reduced the employment-to-population ratio in a local area by 0.4 percentage points, with negative employment effects for all but the most highly educated workers.[33] They also document

33. Using a model of how local economies interact, Acemoglu and Restrepo (2020) then use their regional estimate to calculate that the overall effect on employment-to-population is about 0.2 percentage points or about 400,000 jobs, which are the estimates that they discuss in their conclusion.

that the number of robots in the United States increased by about one per thousand workers from 1993 to 2007. Thus, our baseline estimate of the impact of switching to payroll tax financing on increasing the employment rate of noncollege workers is similar in magnitude to the estimated effect of the increase in robots from 1993 to 2007 on decreasing that employment rate.

To gauge the impact of import exposure, we draw on Autor, Dorn, and Hanson (2013). They estimate that a $1,000 greater import exposure shock (roughly the interquartile range in CZ-level import exposure growth from 2000 through 2007) decreased the noncollege employment rate by 1.11 percentage points. Thus our baseline estimate of the impact of switching to payroll tax financing on increasing the employment rate of noncollege workers is around half the size of the decrease in this employment rate caused by a $1,000 import exposure shock.

V. Labor Market Effects of Rising Health Care Spending

We can also use our framework to consider how the rapidly increasing cost of health insurance in the United States affected labor market outcomes under the current head tax regime. Health expenditures increased in the United States from 6.25 to 16.77 percent of GDP between 1977 and 2019. During this time, the United States "advanced" from being on the higher end of health care spending across countries to being an extreme outlier (panel A of figure 5).

We consider two alternative counterfactual paths for the growth in health care spending: a "no-growth" counterfactual in which US health care spending (and hence the cost of employer-provided health insurance) remains fixed at the 1977 level in real terms and a "Canada" counterfactual in which health care spending in the United States in 2019 was the same share of GDP as it was in Canada (i.e., 10.84 percent of GDP instead of 16.77 percent). In 1977, average employer-provided premiums were $2,760 in 2019 dollars. Under our baseline assumption that only 66 percent of full-time, full-year workers are policyholders, the implied head tax $\tau_{1977} = \$1{,}820 = 0.66 \cdot \$2{,}760$. Thus, compared to our baseline 2019 head tax $\tau_{2019} = \$7{,}758$, under the no-growth counterfactual the head tax would be $5,938 lower than under the observed baseline. Under the Canada counterfactual, we scale our baseline 2019 head tax by the ratio of the Canadian share to the US share of the economy in terms of health care spending. As a result, the counterfactual 2019 head tax is 65 percent (= 10.84/16.77) of our baseline value, or $5,017. Thus, under the Canada counterfactual, the head tax would be $2,741 lower than under the observed baseline.

As discussed in section II, employment outcomes under counterfactual levels of US health care spending and health insurance premiums depend on the workers' amenity value α_g of employer-provided health insurance relative to cash. Conceptually, as discussed in section II, α may range from zero to a value above one and may also vary by group.[34] In practice, there is considerable disagreement about the amenity value of health care spending and further ambiguity about the amenity value of the portion of health care spending that rose over the last four decades. Some researchers emphasize the valuable and health-increasing technological improvements in medicine behind the rise in health spending (Cutler 2022), and others emphasize the large amounts of waste in US health spending (Garber and Skinner 2008). Naturally, these views are not mutually exclusive, and their relative importance likely varies over time and across locations. Absent much guidance from the empirical literature, we present results for four different amenity values: 0, 0.75, 1, and 1.25. In each case, we assume that the amenity value is the same for both groups of workers so that the marginal rate of substitution between health insurance and wages is equal across groups at the margin.

Table 7 presents the results. We estimate that college and noncollege wages would both be about $6,000 higher under the no-growth counterfactual and about $2,750 higher under the Canada counterfactual. These effects are largely insensitive to our assumption about the amenity value. Both groups experience a similar absolute increase in wages when health care spending grows less rapidly.[35] However, this increase is larger proportionally for noncollege workers (12 percent and 5.4 percent, respectively) than for college workers (6 percent and 3 percent, respectively; see table 1, panel A). As a result, the college wage premium is about 10.5 percent lower in the no-growth counterfactual and about 5 percent lower in the Canada counterfactual.

By contrast, the impacts of alternative levels of health care costs on employment vary greatly depending on the assumed amenity value. If workers are indifferent between wages and spending on employer-provided

34. For example, many people might not value medical spending focused on rare conditions.

35. In all cases, we can see that the cost reduction leads to a nearly equal increase in wages for both groups of workers. With linear production (as in our stylized model in section II and $\alpha = 1$) it is straightforward to see that the increase in wages would be the same for both groups (and equal to the decrease in health care costs τ). This is also the case with CES production (see columns 3 and 7), but with CES production and other values of α equilibrium changes in employment that affect marginal productivity of labor have a second-order effect on equilibrium wages that can differ across groups.

Table 7. 2019 Labor Market Effects of Health Care Spending Counterfactuals

	No-growth counterfactual				*Canada counterfactual*			
	$\alpha = 0$	$\alpha = 0.75$	$\alpha = 1$	$\alpha = 1.25$	$\alpha = 0$	$\alpha = 0.75$	$\alpha = 1$	$\alpha = 1.25$
Change in cost, τ	–$5,937	–$5,937	–$5,937	–$5,937	–$2,740	–$2,740	–$2,740	–$2,740
Wages								
Change in college wage, $\Delta(w_C)$	$6,062	$5,969	$5,937	$5,905	$2,799	$2,755	$2,740	$2,726
Change in noncollege wage, $\Delta(w_N)$	$5,840	$5,912	$5,937	$5,962	$2,695	$2,729	$2,740	$2,752
Change in college wage premium (%)	–9.99	–10.43	–10.58	–10.73	–4.88	–5.10	–5.18	–5.25
Employment rate								
Change in college employment rate, $\Delta(P_C)$	1.91 pp	0.48 pp	0.00 pp	–0.48 pp	0.88 pp	0.22 pp	0.00 pp	–0.22 pp
Change in noncollege employment rate, $\Delta(P_N)$	1.84 pp	0.46 pp	–0.00 pp	–0.46 pp	0.85 pp	0.21 pp	–0.00 pp	–0.21 pp
Change in college share of wage bill, $\Delta\left(\frac{w_C L_C}{w_N L_N + w_C L_C}\right)$	–1.30 pp	–1.27 pp	–1.26 pp	–1.25 pp	–0.62 pp	–0.61 pp	–0.61 pp	–0.60 pp

Source: Authors' calculations.

Note: The college wage premium is defined as $CWP = w_C/w_N - 1$. In this table, the percentage change in the college wage premium is calculated as the percentage change in this value when moving from the head tax to the payroll tax, which would be equal to $(CWP^P - CWP^H)/CWP^H$.

health insurance, as represented by a value of α equal to one, then lower spending increases have no employment effects (Summers 1989; Gruber 1994). For α less than one, workers value an increase in wages more than an increase in spending on employer-provided health insurance. As a result, an increase in health care spending is partially a tax on workers, and the incidence of this tax is split between wages and employment. Thus, under counterfactuals of lower spending increases, employment increases. Conversely, for α greater than one, workers value increased spending on employer-provided health insurance more than the increase in wages and so employment increases as health care spending rises. Thus, under counterfactuals of lower spending increases, employment falls.

VI. Discussion

This paper calibrates the effect of the US health care head tax on labor market inequality. We find that if employer-provided health insurance were instead financed by a proportional payroll tax, the college premium would be 11 percent lower, and noncollege employment would be nearly 500,000 higher. Over the last four decades, the rise in the college premium would have been about 20 percent lower and the rise in the employment rate of noncollege adults would have been 4.6 percent higher. These effects are comparable in magnitude to previous estimates of the impact of other leading sources of labor market inequality, including outsourcing, robot adoption, rising trade, and declining unionization. Some of these forces might be driven in part by firms' responses to rising health costs for domestic workers, so they are not mutually exclusive from the health care financing mechanism we focus on here.

Our analyses rely on several important simplifying assumptions. We briefly discuss them here, in the hopes that they may offer fruitful directions for future research.

Perhaps most importantly, our analysis occurs in partial equilibrium. In particular, we have not considered potential effects on employer behavior. In practice, health insurance financing reforms may change whether employers offer health insurance as well as the generosity of plan offerings. There may also be important labor market margins of response, such as the share of workers in part-time versus full-time employment, the propensity of employers to use contract workers, domestic outsourcing, and offshoring. These potential responses are ignored in our counterfactual analysis of the impact of rising US health care spending under current head tax financing, and they are explicitly assumed away in our analysis of a counterfactual,

nationwide payroll tax to finance the health insurance of workers. Naturally, a policy that applies based on workers' current form of health insurance is not practically feasible. As noted at the beginning of the paper, in practice a nationwide payroll tax would need to also finance health insurance for currently uninsured individuals.

Within this partial-equilibrium framework, there are other, potentially important simplifying assumptions. First, like much of the existing literature on labor market inequality, we focus on outcomes such as the college wage premium and noncollege employment and wages. Other measures of inequality such as inequality in total compensation warrant greater attention, especially in light of work on the role of health benefits in affecting a broad measure of compensation inequality (Pierce 2001, 2010; Gittleman and Pierce 2015; Piketty, Saez, and Zucman 2018; Ouimet and Tate 2023). In particular, this work has highlighted that considering health benefits materially reduces measured inequality between high- and low-income workers, while adding incentive compensation and retirement benefits has an opposite effect.

Second, our model of the labor market involves a number of potentially important simplifying assumptions. We consider a textbook, competitive model of the labor market; imperfect competition and other labor market frictions are attracting increasing attention among labor economists (Card 2022), and their quantitative impact on our calibrations is an important and open question. We also abstract from the fact that some full-time, full-year workers are not offered employer-provided health insurance; we instead model the incomplete coverage as happening only through the take-up margin. In practice, firms choose whether or not to offer health insurance in equilibrium (Dey and Flinn 2005), and noncollege workers are disproportionately located in firms that do not offer health insurance (table 1). This sorting—which is itself an equilibrium outcome (Aizawa 2019)—may be partly a reaction to the head tax financing of insurance and likely cushions its impact on labor market inequality. How quantitatively important such worker sorting across firms is for our analysis, however, remains unclear, especially because high-wage workers tend to disproportionately sort into high-wage firms (Song and others 2019).

Expanding our model to allow workers to choose between firms that do and do not offer health insurance would also allow analysis of additional impacts of the financing of employer-provided health insurance on the distribution of employment and earnings. In particular, our analysis follows Case and Deaton (2020b) in focusing on college workers compared to noncollege workers. Others have focused on the decline in the share of workers

in the middle of the wage distribution (Autor 2018; Autor and Dorn 2013). It is conceivable that this "hollowing out" phenomenon may also be affected by the head tax financing of employer-provided health insurance. The lowest wage group may be predominantly covered by Medicaid and in jobs that do not offer health insurance. Accordingly, the middle group would be the one squeezed out by head tax financing, which makes them less attractive relative to both higher-wage workers for whom the head tax is a smaller share of labor costs and lower-wage workers who do not require the head tax to be paid by their employers. Relatedly, it is also conceivable that financing health insurance through the firm contributes to the rise of alternative work arrangements that don't provide insurance and to the fissuring of the workforce (Weil 2014; Katz and Krueger 2017; Card, Heining, and Kline 2013; Song and others 2019).

One important way to make progress on these issues would be through empirical work that identifies exogenous variation in the costs of employer-provided health insurance that can be used to study its impact on labor market outcomes. Importantly, the ideal variation would occur at the labor market rather than firm level, so that it might be possible to estimate impacts on the labor market equilibrium.

Despite all these important directions for further work, our calibrated analysis suggests that the uniquely American approach to financing health insurance could have a quantitatively important impact on labor market inequality. Our analysis suggests that if the cost of health care in the United States continues its rapid rise over the coming years, labor market inequality will also continue to grow in the absence of substantial reforms to how we finance health insurance in America.

ACKNOWLEDGMENTS We are grateful to David Autor, Anne Case, Raj Chetty, David Cutler, Janet Currie, Angus Deaton, Janice Eberly, Jon Gruber, Larry Katz, Pat Kline, Ilyana Kuziemko, and Nathan Hendren for helpful comments, and to Emily Bjorkman, Drew Burd, Patrick Collard, Coly Elhai, Dean Li, and Dustin Swonder for excellent research assistance. We gratefully acknowledge support from the Laura and John Arnold Foundation. Zidar also thanks the NSF for support under grant number 1752431, and Zwick also thanks the Fama Research Fund, Neubauer Family Foundation, the Polsky Center, and the Hultquist Faculty Research Endowment at Chicago Booth.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67, no. 2: 251–333.

Acemoglu, Daron, and David Autor. 2001. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics, Volume 4B*, edited by Orley Ashenfelter and David Card. Amsterdam: North-Holland.

Acemoglu, Daron, and Pascual Restrepo. 2020. "Robots and Jobs: Evidence from U.S. Labor Markets." *Journal of Political Economy* 128, no. 6: 2188–244.

Aizawa, Naoki. 2019. "Labor Market Sorting and Health Insurance System Design." *Quantitative Economics* 10, no. 4: 1401–51.

Aizawa, Naoki, and Hanming Fang. 2020. "Equilibrium Labor Market Search and Health Insurance Reform." *Journal of Political Economy* 128, no. 11: 4258–336.

Autor, David H. 2018. "Trade and Labor Markets: Lessons from China's Rise." Article 431. IZA World of Labor. https://wol.iza.org/articles/trade-and-labor-makets-lessons-from-chinas-rise/long.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103, no. 5: 1553–97.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103, no. 6: 2121–68.

Autor, David H., Claudia Goldin, and Lawrence F. Katz. 2020. "Extending the Race between Education and Technology." *American Economic Association Papers and Proceedings* 110:347–51.

Bagley, Nicholas, Amitabh Chandra, and Austin Frakt. 2015. "Correcting Signals for Innovation in Health Care." Washington: The Hamilton Project, Brookings Institution.

Baily, Martin Neil. 1978. "Some Aspects of Optimal Unemployment Insurance." *Journal of Public Economics* 10, no. 3: 379–402.

Beemon, Dan. 2021. "Job Search, Wage Dispersion, and Universal Healthcare." Working Paper. https://users.ssc.wisc.edu/~beemon/HI_Firm_Response.

Binder, Ariel J., and John Bound. 2019. "The Declining Labor Market Prospects of Less-Educated Men." *Journal of Economic Perspectives* 33, no. 2: 163–90.

Blewett, Lynn A., Julia A. Rivera Drew, Risa Griffin, and Kari C. W. Williams. 2019. *IPUMS Health Surveys: Medical Expenditure Panel Survey, Version 1.1* [data set]. Minneapolis: University of Minnesota. https://doi.org/10.18128/D071.V1.1.

Blumenthal, David. 2006. "Employer-Sponsored Health Insurance in the United States—Origins and Implications." *New England Journal of Medicine* 355:82–88.

Blundell, Richard, and Ben Etheridge. 2010. "Consumption, Income and Earnings Inequality in Britain." *Review of Economic Dynamics* 13, no. 1: 76–102.

Bound, John, and George Johnson. 1992. "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations." *American Economic Review* 82, no. 3: 371–92.

Brzozowski, Matthew, Martin Gervais, Paul Klein, and Michio Suzuki. 2010. Consumption, Income, and Wealth Inequality in Canada [data set]. *Review of Economic Dynamics* 13, no. 1: 52–75.

Card, David. 2022. "Who Set Your Wage?" *American Economic Review* 112, no. 4: 1075–90.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *Quarterly Journal of Economics* 128, no. 3: 967–1015.

Card, David, Thomas Lemieux, and W. Craig Riddell. 2004. "Unions and Wage Inequality." *Journal of Labor Research* 25:519–59.

Case, Anne, and Angus Deaton. 2020a. "America Can Afford a World-Class Health System. Why Don't We Have One?" *New York Times*, April 14. https://www.nytimes.com/2020/04/14/opinion/sunday/covid-inequality-health-care.html.

Case, Anne, and Angus Deaton. 2020b. *Deaths of Despair and the Future of Capitalism*. Princeton, N.J.: Princeton University Press.

Chetty, Raj. 2006. "A New Method of Estimating Risk Aversion." *American Economic Review* 96, no. 5: 1821–34.

Chetty, Raj. 2012. "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica* 80, no. 3: 969–1018.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2013. "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-analysis of Extensive Margin Elasticities." *NBER Macroeconomics Annual* 27, no. 1: 1–56.

Chetverikov, Denis, Bradley Larsen, and Christopher Palmer. 2016. "IV Quantile Regression for Group-Level Treatments, with an Application to the Distributional Effects of Trade." *Econometrica* 84, no. 2: 809–33.

Congressional Budget Office. 2019. *Federal Subsidies for Health Insurance Coverage for People under Age 65: 2019 to 2029*. Washington: Congressional Budget Office.

Cutler, David M. 2022. *Your Money or Your Life: Strong Medicine for America's Health Care System*. New York: Oxford University Press.

Cutler, David M., and Brigitte C. Madrian. 1998. "Labor Market Responses to Rising Health Insurance Costs: Evidence on Hours Worked." *RAND Journal of Economics* 29, no. 3: 509–30.

Dey, Matthew S., and Christopher J. Flinn. 2005. "An Equilibrium Model of Health Insurance Provision and Wage Determination." *Econometrica* 73, no. 2: 571–627.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach." *Econometrica* 64, no. 5: 1001–44.

Domeij, David, and Martin Flodén. 2010. "Inequality Trends in Sweden 1978–2004." *Review of Economic Dynamics* 13, no. 1: 179–208.

El-Sibaie, Amir. 2018. "2019 Tax Brackets." Tax Foundation, November 28. www.taxfoundation.org/2019-tax-brackets/.

Eurostat. 2022. EU Statistics on Income and Living Conditions [data set]. ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions.

Fang, Hanming, and Dirk Krueger. 2021. "The Affordable Care Act after a Decade: Its Impact on the Labor Market and the Macro Economy." Working Paper 29240. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29240.

Farber, Henry S., Daniel Herbst, Ilyana Kuziemko, and Suresh Naidu. 2021. "Unions and Inequality over the Twentieth Century: New Evidence from Survey Data." *Quarterly Journal of Economics* 136, no. 3: 1325–85.

Feenstra, Robert C., and Gordon H. Hanson. 2003. "Global production sharing and rising inequality: A survey of trade and wages." In *Handbook of International Trade*, edited by E. Kwan Choi and James Harrigan. Malden, Mass.: Blackwell.

Feldstein, Martin S. 1973. "The Welfare Loss of Excess Health Insurance." *Journal of Political Economy* 81, no. 2: 251–80.

Feldstein, Martin S., and Bernard Friedman. 1977. "Tax Subsidies, the Rational Demand for Insurance and the Health Care Crisis." *Journal of Public Economics* 7, no. 2: 155–78.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, and Michael Westberry. 2021. *Integrated Public Use Microdata Series, Current Population Survey: Version 9.0* [data set]. Minneapolis: University of Minnesota.

Fuchs-Schuendeln, Nicola, Dirk Krueger, and Mathias Sommer. 2010. Inequality Trends for Germany in the Last Two Decades: A Tale of Two Countries [data set]. *Review of Economic Dynamics* 13, no. 1: 103–32.

Garber, Alan M., and Jonathan Skinner. 2008. "Is American Health Care Uniquely Inefficient?" *Journal of Economic Perspectives* 22, no. 4: 27–50.

Garthwaite, Craig, Tal Gross, and Matthew J. Notowidigdo. 2014. "Public Health Insurance, Labor Supply, and Employment Lock." *Quarterly Journal of Economics* 129, no. 2: 653–96.

Gittleman, Maury, and Brooks Pierce. 2015. "Pay for Performance and Compensation Inequality: Evidence from the ECEC." *ILR Review* 68, no. 1: 28–52.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race between Education and Technology.* Cambridge, Mass.: Harvard University Press.

Goldschmidt, Deborah, and Johannes F. Schmieder. 2017. "The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure." *Quarterly Journal of Economics* 132, no. 3: 1165–217.

Gregware, Jim. 2017. "National Survey of Employer-Sponsored Health Plans 2016" (slides). New York: Mercer.

Gruber, Jonathan. 1994. "The Incidence of Mandated Maternity Benefits." *American Economic Review* 84, no. 3: 622–41.

Gruber, Jonathan. 2000. "Health Insurance and the Labor Market." In *Handbook of Health Economics Volume 1A*, edited by Anthony J. Culyer and Joseph P. Newhouse. Amsterdam: North-Holland.

Gruber, Jonathan. 2002. "Taxes and Health Insurance." *Tax Policy and the Economy* 16:37–66.

Gruber, Jonathan. 2008. "Covering the Uninsured in the United States." *Journal of Economic Literature* 46, no. 3: 571–606.

Gruber, Jonathan. 2011. "The Tax Exclusion for Employer-Sponsored Health Insurance." *National Tax Journal* 64, no. 2.2: 511–30.

Gruber, Jonathan, and Brigitte C. Madrian. 2004. "Health Insurance, Labor Supply, and Job Mobility: A Critical Review of the Literature." In *Health Policy and the Uninsured*, edited by Catherine McLaughlin. Washington: Urban Institute Press.

Gruber, Jonathan, and Robin McKnight. 2003. "Why Did Employee Health Insurance Contributions Rise?" *Journal of Health Economics* 22, no. 6: 1085–104.

Gruber, Jonathan, and Kosali Simon. 2008. "Crowd-out 10 Years Later: Have Recent Public Insurance Expansions Crowded Out Private Health Insurance?" *Journal of Health Economics* 27, no. 2: 201–17.

Jappelli, Tullio, and Luigi Pistaferri. 2010. Code and Data Files for "Does Consumption Inequality Track Income Inequality in Italy?" *Review of Economic Dynamics* 13, no. 1: 133–53.

Katz, Lawrence F. 2008. Comment on "Trade and Wages, Reconsidered," by Paul R. Krugman. *Brookings Papers on Economic Activity*, Spring, 143–49.

Katz, Lawrence F., and Alan B. Krueger. 2017. "The Role of Unemployment in the Rise in Alternative Work Arrangements." *American Economic Review* 107, no. 5: 388–92.

Katz, Lawrence F., and Kevin W. Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107, no. 1: 35–78.

Keane, Michael, and Richard Rogerson. 2012. "Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom." *Journal of Economic Literature* 50, no. 2: 464–76.

Krueger, Dirk, Fabrizio Perri, Luigi Pistaferri, and Giovanni L. Violante. 2010. "Cross-Sectional Facts for Macroeconomists." *Review of Economic Dynamics* 13, no. 1: 1–24.

Krugman, Paul R. 2008. "Trade and Wages, Reconsidered." *Brookings Papers on Economic Activity*, Spring, 103–37.

Lee, David S. 1999. "Wage Inequality in the United States during the 1980s: Rising Dispersion or Falling Minimum Wage?" *Quarterly Journal of Economics* 114, no. 3: 977–1023.

Lurie, Ithai Z., and Corbin L. Miller. 2023. "Employer-Sponsored Health Insurance Premiums and Income in U.S. Tax Data." *Journal of Public Economics* 224:104942.

Mankiw, N. Gregory, and Lawrence H. Summers. 2015. "Uniting behind the Divisive 'Cadillac' Tax on Health Plans." *New York Times*, October 24. https://www.nytimes.com/2015/10/25/upshot/uniting-behind-the-divisive-cadillac-tax-on-health-plans.html.

Mui, Preston, and Benjamin Schoefer. 2021. "Reservation Raises: The Aggregate Labor Supply Curve at the Extensive Margin." Working Paper 28770. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28770.

Organisation for Economic Co-operation and Development (OECD). 2019. *Health at a Glance 2019: OECD Indicators*. Paris: OECD Publishing. https://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-2019_4dd50c09-en.

Organisation for Economic Co-operation and Development (OECD). 2022. *Education at a Glance 2022: OECD Indicators.* Paris: OECD Publishing. https://www.oecd.org/education/EAG2022_X3.pdf.

Ouimet, Paige, and Geoffrey A. Tate. 2023. "Firms with Benefits? Nonwage Compensation and Implications for Firms and Labor Markets." Working Paper 31463. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w31463.

Pauly, Mark V. 1986. "Taxation, Health Insurance, and Market Failure in the Medical Economy." *Journal of Economic Literature* 24, no. 2: 629–75.

Pierce, Brooks. 2001. "Compensation Inequality." *Quarterly Journal of Economics* 116, no. 4: 1493–525.

Pierce, Brooks. 2010. "Recent Trends in Compensation Inequality." In *Labor in the New Economy*, edited by Katharine G. Abraham, James R. Spletzer, and Michael J. Harper. Chicago: University of Chicago Press.

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.

Robertson, Christopher. 2014. "Scaling Cost-Sharing to Wages: How Employers Can Reduce Health Spending and Provide Greater Economic Security." *Yale Journal of Health Policy, Law, and Ethics* 14, no. 2: 239–95.

Saez, Emmanuel, and Gabriel Zucman. 2019a. *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*. New York: W. W. Norton.

Saez, Emmanuel, and Gabriel Zucman. 2019b. "We Can Afford Medicare for All . . . and It Could Even Deliver a Huge Pay Raise to the Middle Class." Politico, November 25. https://www.politico.com/news/agenda/2019/11/25/agenda-can-we-afford-medicare-for-all-071560.

Sammer, Joanne. 2017. "Tying Health Plan Premiums to Salary Can Aid Lower-Paid Earners." Society for Human Resource Management, October 19.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "Capitalists in the Twenty-First Century." *Quarterly Journal of Economics* 134, no. 4: 1675–745.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2022. "The Rise of Pass-Throughs and the Decline of the Labor Share." *American Economic Review: Insights* 4, no. 3: 323–40.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2019. "Firming Up Inequality." *Quarterly Journal of Economics* 134, no. 1: 1–50.

Starr, Paul. 1982. *The Social Transformation of American Medicine: The Rise of a Sovereign Profession and the Making of a Vast Industry*. New York: Basic Books.

Summers, Lawrence H. 1989. "Some Simple Economics of Mandated Benefits." *American Economic Association Papers and Proceedings* 79, no. 2: 177–83.

Thomasson, Melissa A. 2002. "From Sickness to Health: The Twentieth-Century Development of U.S. Health Insurance." *Explorations in Economic History* 39, no. 3: 233–53.

Thomasson, Melissa A. 2003. "The Importance of Group Coverage: How Tax Policy Shaped U.S. Health Insurance." *American Economic Review* 93, no. 4: 1373–84.

Vogel, Jonathan. 2022. "The Race between Education, Technology, and Institutions." Working Paper 30311. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30311.

Weil, David. 2014. *The Fissured Workplace: Why Work Became So Bad for So Many and What Can Be Done to Improve It*. Cambridge, Mass.: Harvard University Press.

Comments and Discussion

COMMENT BY

DAVID CUTLER Among high-income countries, the United States is unusual in relying so heavily on employer-based financing of health insurance. Most countries finance health insurance through an income or payroll tax, as part of a national health insurance system. Such a system assesses a fee more than actuarial cost of insurance on higher-income people and less than actuarial cost of insurance on low-income people. In contrast, a premium-based system charges the same dollar amount to everyone—the classic head tax.

Because health insurance is so expensive, the implicit cost of health insurance for each employee differs greatly between a head tax and a payroll tax. An average family health insurance policy costs around $20,000 per year. If one considers an employer paying for that entire amount, that is 40 percent of the wage for a $50,000 per year employee, but only 10 percent of the wage for a $200,000 per year employee. If the employer instead paid a fixed share of payroll for insurance, the required tax rate would be 16 percent (in this example), and the cost of health insurance would fall to $8,000 annually for the low-wage employee and rise to $32,000 annually for the high-wage worker.

A key question is what impact such a change, $12,000 per employee per year in this example, would have on earnings and employment of different workers. Finkelstein, McQuillan, Zidar, and Zwick explore that issue in this paper. They use a model of the labor market to simulate the impact of moving from a head tax to payroll tax on earnings and employment of workers with more and less education. The theoretical effect is clear. With the lower cost for workers with less education, labor demand for such

Brookings Papers on Economic Activity, Spring 2023: 476–503 © 2024 The Brookings Institution.

workers will rise. This will lead to some combination of increased earnings and greater employment, with the relative split between the two determined by the labor supply elasticity. The same is true in the opposite direction for workers with more education.

The authors' results suggest that the impact will be largely on wages, though employment will be somewhat affected as well. If health insurance were financed by a payroll tax instead of a head tax, wages for noncollege-educated workers would rise by roughly $1,700 per year, and the college wage premium would fall by 11 percent. Employment of noncollege-educated workers would increase by nearly 500,000 people. There would be a similar-sized reduction in employment of college-educated workers, of roughly 400,000 workers. The net impact would be a modest increase in employment, roughly 87,000 workers, and a tilting of earnings toward workers with less education. The relatively greater impact of the change on wages relative to employment is a reflection of the fact that the labor supply elasticities the authors use are not particularly large.

The authors do a similar calculation about the impact of the increase in health insurance costs over time. Here, the analysis is more complex, because the additional spending on health insurance is buying additional services (in part), which may themselves influence labor supply. Using an assumption of no benefit from additional spending, so that the increase in premiums is entirely a tax, they find that workers with both more and less education would earn more and work more hours had health insurance cost not increased. This is understandable since insurance costs for both groups of workers would fall. The overall income gains would be greater for the noncollege-educated group, as one expects given the higher cost burden for that population.

It is important to note that payroll tax financing cannot be administered in the same fashion as head tax financing. A head tax can be implemented in private markets. Insurers can, and some do, charge firms a fixed amount per insured individual, equal to expected medical costs plus overhead. At such a premium, the insurer will break even for each firm. In contrast, the payroll tax has to be administered through the public sector. In the example above, if the firm hired an additional low-wage worker, a constant payroll tax rate would not cover expected health insurance costs for the firm. Thus, the privately set tax rate would have to increase or decline with the wage composition of employment. This adjustment would have to be sufficient to raise the total amount paid by the firm by $20,000 for each new worker hired, independent of the worker's wage. A savvy firm would know that the cost of each worker was the full cost of insurance for an additional person,

independent of how the premium was quoted. For this reason, a private payroll tax financing system would not be feasible, and the change that the authors consider, from a head tax to a payroll tax, is not a trivial one.

One implication of the authors' analysis is that wage inequality is not the best measure of labor market differences across groups. Rather, the right measure of inequality is compensation inequality. Wages are one part of total compensation, but there are other parts as well.

The authors do not analyze trends in total compensation in the United States over time or across countries. However, others have. Most notable is research by Pierce (2010). Pierce compares changes in compensation in different parts of the compensation distribution to changes in wages in different parts of the wage distribution. Overall, the growth of compensation inequality between 1987 and 2007 is roughly the same as the growth of wage inequality. This seems unusual given the very rapid growth of health insurance over this time, which raises compensation the most for low- and middle-income employees. The reason is that there is an offsetting effect: at the same time that health insurance costs were rising, other benefits increased more for high-wage workers, such as pensions and bonuses or incentive pay. On net, the two roughly cancel out.

The importance of differential changes in benefits other than health insurance suggests a key issue with the incidence analysis in the paper. It may be that the impact of rising health insurance premiums is not on wages but rather on other benefits. For example, firms may respond to rising health insurance costs by moving money out of wages and salaries and into pensions or bonus arrangements, especially those targeted to high-income workers. If this were the impact, changing the financing of health insurance in the United States would not necessarily change wages but might instead change other benefits.

The net impact of an offset to benefits rather than wages depends on the marginal valuation of benefits and wages. If all benefits are valued dollar for dollar at the margin, employees would not care where the offset occurred. Of course, there are different tax implications of benefits and wages or salaries, which the government would need to consider. And the impact on well-being would be very different if not all employees value benefits dollar for dollar at the margin. For all these reasons, understanding whether health insurance costs have led to reduced wages and salaries or other benefits is a very important issue.

This discussion raises a broader point about the incidence of the current health insurance financing system, as well as proposed alternatives. To do

their analysis, the authors need to make assumptions about incidence—How does a firm respond when faced with health insurance costs that vary in how they are assessed? The authors are upfront about the need to do this and are also clear in the assumptions they make. They treat the current financing system as a head tax, and the alternative as a tax per dollar of earnings. Viewing the payroll tax this way is reasonably standard. Since the payroll tax rate is specified in advance, and it is not complicated for firms to understand, it makes sense to assume that firms treat the payroll tax as a tax per dollar of wage and salary payment.

In contrast, treating the current financing system as leading to a reduction in the wage per hour worked is a more complicated assumption. There is not a great deal of evidence about the incidence of current health insurance payments. Some studies suggest a dollar for dollar trade-off between health insurance costs and wages (Gruber 1994; Sheiner 1999), though not all do so (Levy and Feldman 2001). However, full impact on wage rates is not the only theoretical prediction. Provided firms can adjust hours as well as wages, one would not expect health insurance costs to affect only wages (Gruber 2000). Rather, firms should have employees work more hours and hire fewer workers in total. The additional hours may come at a greater marginal cost, but the reduction in employment it leads to would reduce the overall compensation bill. Along these lines, some studies have found impacts of health insurance costs on hours of work (Cutler and Madrian 1998), though work on this topic is not fully complete.

The issue of how high insurance costs affect the full employment contract is very important because understanding how firms treat the current financing system will directly affect the implications of moving to a new financing system. In particular, if firms respond to a health insurance bill presented by an insurer by reducing the wages of all employees by the same percentage (not the same dollar amount), this will be the same incidence as the payroll tax alternative that the authors consider.

Three features of the world suggest to me that the incidence of current health insurance financing is more complex than what the authors assume. The first fact is that there are large firm effects in compensation. The same worker, with the same occupation, earns different amounts depending on which firm they work for (Card and others 2018). In essence, the administrative staff at a fancy software engineering firm earns more than the same administrative staff at a low-wage retail establishment. It may be that this is irrelevant with respect to the incidence of health insurance—the high-wage firm may still offset higher insurance costs by reducing the wage rate

dollar for dollar—but it may also be relevant. The same firms that offer above-average wages for similar workers may not offset the higher costs of health insurance to those workers in the same fashion.

The second observation is that firms seem to care a lot about health insurance costs, even when medical spending trends are affecting all firms roughly equally. Costs that affect one firm and not others will necessarily affect that firm's decision-making. If one firm has less healthy employees than another and thus higher medical spending than its competitor, the firm with higher medical costs will have to decide how to adjust—reduce wages, earn lower profits, or some other method. However, if all firms are affected by rising costs, the equilibrium would be for all wages to fall dollar for dollar. In this equilibrium, firms would not care about the cost of health care. They would simply take the total cost of health insurance as given and set wages based on them.

In reality, firms seem to care greatly about how much they are paying for health insurance, even when they pay more because of general cost increases. In a recent survey, roughly one-third of firms expressed dissatisfaction with the overall cost of medical care (Claxton and others 2022). The vast majority of the remainder are "satisfied" but not "very satisfied." Such a situation would not be apparent if firms took health costs as given and costs were rising uniformly for all firms.

Third, firms seem to respond to high health costs not just by reducing wages but by changing the structure of their health insurance in a way that reduces the costs they face. Perhaps the biggest change in health insurance in recent years has been the shift to high deductible plans. In the 1990s, less than 10 percent of workers with employer-sponsored insurance were enrolled in high deductible health plans. Today, it is one-third of workers (Claxton and others 2022). There are many reasons for this change; the rising cost of medical care is one of them (Chernew, Cutler, and Keenan 2005). As medical costs have increased, firms have responded by reducing the generosity of what they offer.

For many reasons, this is not what theory would suggest. One major reason is the tax exclusion of employer-provided health insurance. Employer payments for health insurance are not taxed as income, where wages and salaries are. Thus, from an employee's standpoint, it is best to run any medical spending which is sure to occur through health insurance. In addition, rising cost sharing exposes employees to more risk. This risk bearing is a welfare reduction, and employees might be willing to pay for (i.e., in lower wages) a more generous package than is the norm in a high deductible policy. Thus, something seems off with the basic theory.

In thinking about alternative models of wage setting in the presence of high and rising health insurance costs, there are several directions one can go. One traditional way is by building in nominal wage stickiness. Offsetting rising health insurance costs with lower wages may not be feasible if it involves reductions in nominal wages, as opposed to lower wage increases. In the presence of nominal wage stickiness, the offset to health insurance costs may occur more in employment and less than wages.

A second type of model considers separation of benefits costs and wages. Imagine a model of the following form. First, firms pay for health and other pre-committed benefits out of revenues. Second, employees outside of senior executives are paid a wage based on opportunity cost. Third, residual revenues are distributed to owners, either as bonuses, performance pay, or in some other nonwage form. In the classic version of wage setting, employees recognize the value of health benefits at the firm and are willing to take a wage discount for better benefits. However, it is possible that employees have difficulty comparing health benefits at their current firm with other potential employers. They may assume health benefits are "good" everywhere, or at minimum very difficult to compare, and thus only consider the wage on the job relative to the wage on other potential jobs. In this case, wages could not be fully adjusted for increasing health costs or workers would leave the firm.

In such a setting, high benefits costs will primarily come out of executive compensation—the residual after wages or salaries and benefits for nonexecutives are paid. Employment may or may not be affected, depending on the elasticity of demand for nonexecutive labor. This model would have a very different application for incidence than the one that the authors consider.

A third type of model involves differential valuation of health insurance by different types of people. It is easiest to introduce the model starting with the pattern of wage inequality that we observe in the data. Song and others (2019) analyze changes in the wage distribution over time.[1] They highlight three aspects contributing to wage inequality. The first is that high-wage people are more likely to work in high-wage firms over time—superstar computer programmers increasingly work at superstar software engineering companies. My guess is that this has relatively little to do with health insurance and much more to do with geographic mobility by income.

The second component is that there is more segregation of high-wage workers into high-wage firms than used to be the case. An example of this

1. See also Haltiwanger, Hyatt, and Spletzer (2022).

would be a software engineering firm that used to have its own cleaning staff but now contracts for those services from outside companies. The result is that high wages are more concentrated among high-income earners than used to be the case. This may absolutely be a result of rising health costs, which I return to. The third part of rising wage inequality is absolute reductions in wages in what researchers call "mega firms"—firms that employ 10,000 or more workers. There are roughly 750 such firms, including many well-known companies: Walmart, Coca-Cola, the major airlines, and so on (Song and others 2019). Within these firms, low-wage workers have traditionally been paid more than the wage workers in their comparable smaller competitors, as noted above. Over time, wage differentials for low-wage workers in these firms have been eroding. Indeed, wages have been falling in nominal terms in many of these firms. This also might be a direct result of rising health insurance costs.

The theory that I have in mind is easiest to consider in the context of rising medical care costs. In such a setting, it is important to understand why health insurance costs are increasing. The evidence suggests several reasons. The first is the rising cost of new medical technologies. Historically, this is the most common explanation for rising health costs (Newhouse 1992; Cutler, Rosen, and Vijan 2006; Cutler and others 2022).

That explanation remains true, but with a twist that is particularly important. A lot of recent new technologies have been in pharmaceuticals and other treatments that apply to a relatively small group of people, who are ex ante easy to identify. For example, a few decades ago, there was little effective treatment for people with rheumatoid arthritis. Now, there are a number of different possibilities, each with a very high price tag. Treating rheumatoid arthritis can cost upward of $20,000 per year (Cardarelli 2016). This type of technological change has led to a large increase in spending at the high end of the spending distribution. We see this in pharmaceutical data. The Massachusetts Health Policy Commission (2023) found that between 2017 and 2021, the median price of a prescription was essentially unchanged, while the 95th percentile price increased by over 50 percent.

A hallmark of rheumatoid arthritis, and many similar conditions, is that it is extremely persistent: people who have the condition in one year are almost certain to have it the next year, and people who do not have the condition are very unlikely to acquire it. For example, about 1 percent of the US population has rheumatoid arthritis (Xu and Wu 2021), and since there is no cure, people who are treated for it in one year will almost certainly be treated for it in the future. At the same time, the annual onset of new cases is much smaller (Abhishek and others 2017).

Now consider what happens when a new drug is developed to manage rheumatoid arthritis. If a firm that is providing insurance charges everyone more to pay for the cost of the new drug (for example, by reducing wages across the board), a lot of people will be paying more for something they have very little expectation of using. To be sure, they would like actuarially fair insurance to cover rheumatoid arthritis medications, but for someone without rheumatoid arthritis, the actuarially fair policy costs very little. Any policy that covers people with preexisting cases reasonably well will be actuarially unfair for those without the condition. Of course, some people buy insurance even for risks that are low. But given the very low likelihood of developing rheumatoid arthritis for those who do not have it, and the amount it adds to premiums, it is entirely possible that people without rheumatoid arthritis would choose not to purchase coverage for the new medication.

Targeted new treatments are one source of cost increase. There are others that are important as well. A second source of cost increases is higher prices charged by providers, especially providers with market power. A large body of literature shows the providers with market power charge more than those without. For example, nationally ranked hospitals with a reputation for high quality receive extremely high prices relative to those without such a reputation (Cooper and others 2019). Over time, the medical system has consolidated (Cutler and Scott Morton 2013). What used to be many independent hospitals is now a smaller number of health systems, including hospitals and other care providers (Beaulieu and others 2023). These health systems often have the market power necessary to charge higher prices. Indeed, in the past few decades, research shows an important role for rising prices in higher medical care spending (Tollen and Keating 2023).

Paying more for better providers may be worth it, but raising prices may be particularly pernicious for low-income workers. The reason is that the providers charging more are often ones used disproportionately by the highest-wage workers. For example, in Boston, the highest prices are charged by the academic medical centers—the integrated system of Massachusetts General Hospital and Brigham and Women's Hospital (Massachusetts Health Policy Commission 2023). Those hospitals are difficult to access for people who have low and middle incomes, who do not live nearby, and who have difficulty following through on the various steps required to schedule appointments at these institutions. My guess is that the bulk of those with private insurance who use Mass General Brigham tend to be higher than average income. Thus, even if one thought the higher prices at prestigious institutions were justified by higher quality,

a typical insurance policy charges everyone more to pay for care used disproportionately by the best off.

A third explanation for rising medical spending is higher administrative costs. Both insurers and providers incur administrative costs as they haggle over what services can be provided and how much will be paid for them. These administrative costs have increased over time as underlying medical costs have risen. Indeed, high prices are likely causing high administrative spending. To return to the example of rheumatoid arthritis, in order to minimize the cost of new rheumatoid arthritis drugs, insurers may require documentation before approving each use of the medication. Managing such documentation adds to cost, without providing any direct patient benefits.

The effect of these three components—new medications, provider rents, and increased administrative costs—is that a good deal of health insurance cost increases is likely not associated with measurable quality gains. That is particularly true at the margin, where a typical, generally healthy employee may experience relatively little new technology to help improve their condition. If that is the case, the α term in the paper may be far below one, again at the margin (that is, for the rise in medical spending).

In response to a low marginal value of medical spending, firms will find it difficult to offset the cost of health insurance in lower wages. This will lead them to cut back on the generosity of the benefits package. In many instances, this is done by raising the cost sharing that individuals need to make for coverage. For example, medications for rheumatoid arthritis can be put on a higher tier of the formulary where much greater out-of-pocket payments are required, along with the value of prior approval noted above. One limit of this is the high deductible health insurance plan, for which the employee has to pay several thousand dollars before insurance contributes anything. The consequence of these actions is that health insurance premiums stay low, but health insurance becomes a less desirable product.

As the health insurance policy becomes less generous, the overall value of the policy declines, especially for those with low and middle incomes. Think again about a high deductible health insurance plan. Only one-third of American families have money they consider savings (Pew Charitable Trusts 2015), and total financial assets for the median family with $40,000–$60,000 in annual earnings is only $4,000.[2] Effectively, for many families, having a high deductible insurance policy is close to being uninsured. If the

2. Board of Governors of the Federal Reserve System, "Survey of Consumer Finances, 1989–2019," https://www.federalreserve.gov/econres/scf/dataviz/scf/table/#series:Transaction_Accounts;demographic:inccat;population:all;units:median.

family has to pay the first $2,500 before any insurance reimbursement, they may act as if they are uninsured—they may avoid all care to limit the possibility of facing medical bills they cannot afford. That translates into a still lower value of insurance. Employees will be less willing to give up wages for benefits when the benefits realized are not very helpful.

The impacts of such changes are likely to be felt in many ways. As the overall value of health insurance falls, firms will face incentives to break into high- and low-wage subcomponents, outsourcing low-wage workers to companies that do not provide health benefits (the second of the three explanations for increasing wage inequality noted above). Some literature suggests that high health insurance costs may be one of the explanations for outsourcing (Weil 2014). For this reason, I do not view outsourcing of workers as a separate explanation for the impact of health insurance on wage inequality, the way that the authors sometimes set it up. Similarly, the push to replace human labor with machines may also be a consequence of rising health insurance costs. Machines do not need health insurance, and thus wage trade-offs are not necessary.

In the end, I think the authors have identified an important issue associated with wages for low-income workers. However, I suspect the true model of the world runs even deeper than what they propose. The fact that a good deal of health care costs are not associated with technologies or care settings that benefit typical workers, especially generally healthy low- and middle-income workers, means that rising medical costs may have far-reaching implications for the structure of employment and wages. The economic importance of bringing rationality to medical care may be even bigger than is generally discussed.

REFERENCES FOR THE CUTLER COMMENT

Abhishek, Abhishek, Michael Doherty, Chang-Fu Kuo, Christian D. Mallen, Weiya Zhang, and Matthew J. Grainge. 2017. "Rheumatoid Arthritis Is Getting Less Frequent—Results of a Nationwide Population-Based Cohort Study." *Rheumatology (Oxford)* 56, no. 5: 736–44.

Beaulieu, Nancy D., Michael E. Chernew, J. Michael McWilliams, Mary Beth Landrum, Maurice Dalton, Angela Yutong Gu, and others. 2023. "Organization and Performance of US Health Systems." *JAMA* 329, no. 4: 325–35.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics* 36, no. S1: S13–S70.

Cardarelli, William J. 2016. "Managed Care Decision Makers: Understanding the Full Scope of Rheumatoid Arthritis Management." *American Journal of Managed Care* 22, no. 14: S462–69.

Chernew, Michael, David M. Cutler, and Patricia Seliger Keenan. 2005. "Increasing Health Insurance Costs and the Decline in Insurance Coverage." *Health Services Research* 40, no. 4: 1021–39.

Claxton, Gary, Matthew Rae, Emma Wager, Gregory Young, Heidi Whitmore, Jason Kerns, Greg Shmavonian, and Anthony Damico. 2022. *2022 Employer Health Benefits Survey*. San Francisco: Kaiser Family Foundation.

Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John Van Reenen. 2019. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured." *Quarterly Journal of Economics* 134, no. 1: 51–107.

Cutler, David M., Kaushik Ghosh, Kassandra L. Messer, Trivellore Raghunathan, Allison B. Rosen, and Susan T. Stewart. 2022. "A Satellite Account for Health in the United States." *American Economic Review* 112, no. 2: 494–533.

Cutler, David M., and Brigitte Madrian. 1998. "Labor Market Responses to Rising Health Insurance Costs: Evidence on Hours Worked." *RAND Journal of Economics* 29, no. 3: 509–30.

Cutler, David M., Allison B. Rosen, and Sandeep Vijan. 2006. "The Value of Medical Spending in the United States, 1960–2000." *New England Journal of Medicine* 355:920–27.

Cutler, David M., and Fiona Scott Morton. 2013. "Hospitals, Market Share, and Consolidation." *JAMA* 310, no. 18: 1964–70.

Gruber, Jonathan. 1994. "The Incidence of Mandated Maternity Benefits." *American Economic Review* 84, no. 3: 622–41.

Gruber, Jonathan. 2000. "Health Insurance and the Labor Market." In *Handbook of Health Economics Volume 1A*, edited by Anthony J. Culyer and Joseph P. Newhouse. Amsterdam: North-Holland.

Haltiwanger, John C., Henry R. Hyatt, and James R. Spletzer. 2022. "Industries, Mega Firms, and Increasing Inequality." Discussion Paper 15197. Bonn: IZA Institute of Labor Economics. https://docs.iza.org/dp15197.pdf.

Levy, Helen, and Roger Feldman. 2001. "Does the Incidence of Group Health Insurance Fall on Individual Workers?" *International Journal of Health Care Finance and Economics* 1, nos. 3–4: 227–47.

Massachusetts Health Policy Commission. 2023. *Annual Health Care Cost Trends Report and Policy Recommendations*. Boston: Author. https://www.mass.gov/doc/2023-health-care-cost-trends-report/download.

Newhouse, Joseph P. 1992. "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives* 6, no. 3: 3–21.

Pew Charitable Trusts. 2015. *Survey of American Family Finances*. https://www.pewtrusts.org/-/media/assets/2015/03/fsm-poll-results-toplines_artfinal_v3.pdf.

Pierce, Brooks. 2010. "Recent Trends in Compensation Inequality." In *Labor in the New Economy*, edited by Katharine G. Abraham, James R. Spletzer, and Michael J. Harper. Chicago: University of Chicago Press.

Sheiner, Louise. 1999. "Health Care Costs, Wages, and Aging." FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2019. "Firming Up Inequality." *Quarterly Journal of Economics* 134, no. 1: 1–50.

Tollen, Laura, and Elizabeth Keating. 2023. *A Road Map for Action: Recommendations of the Health Affairs Council on Health Care Spending and Value*. Washington: Health Affairs Council on Health Care Spending and Value. https://www.healthaffairs.org/r/healthspending/health-affairs-spending-council-report.pdf.

Weil, David. 2014. *The Fissured Workplace: Why Work Became So Bad for So Many and What Can Be Done to Improve It*. Cambridge, Mass.: Harvard University Press.

Xu, Yingke, and Qing Wu. 2021. "Prevalence Trend and Disparities in Rheumatoid Arthritis among US Adults, 2005–2018." *Journal of Clinical Medicine* 10, no. 15: 3289.

COMMENT BY

ANGUS DEATON[1] I welcome this paper and congratulate its authors. It is an important contribution to the policy debates about the American health care system and how it is financed. I also have a personal reason to welcome it because it takes seriously and extends an argument that Anne Case and I made in our book *Deaths of Despair and the Future of Capitalism* (Case and Deaton 2020).

I begin with an account of the origins of our argument about the health wedge, that the financing of health care lowers both wages and the demand for labor and does so by more for less-skilled workers. In our book, we focused on workers without a four-year college degree, but the argument is more general.

Our work began with a paper (Case and Deaton 2015) in which we identified an unprecedented increase in mortality among midlife white non-Hispanic men and women in the twenty-first century. An analysis of the death certificates showed that the increase was almost entirely confined to men and women without a four-year college degree. An important factor in the increase was an epidemic of suicides, drug overdoses, and deaths from alcoholic liver disease, deaths that Case (2015) subsequently labeled "deaths of despair." In later work (Case and Deaton 2020, 2022), we documented that after 2013 deaths of despair had spread beyond white Americans into the Black and Hispanic communities. Here too, those with a four-year college degree were largely immune.

From the beginning, we suspected that the underlying causes lay in an increasingly difficult labor market for less-skilled workers and the cascading

1. I am grateful for helpful comments and discussions to Bettina Aten, Anne Case, Jason Furman, and Marshall Leinsdorf. They did their best and remaining errors are the responsibility of the author. I acknowledge support from the National Institute of Aging through a grant to the NBER, award number R01AG060104.

effects of the loss of good jobs on community destruction, on morbidity as well as on mortality, on marriages, and on social and institutional attachment. An immediate challenge for such accounts was the fact that deaths of despair were not showing up in other countries, at least not in anything like the same rate as in the United States. Similar, but much smaller, epidemics existed in England and in other English-speaking countries but, with the notable exception of Scotland, the United States stood out among rich countries. Given this, stories of globalization and automation, despite the evidence from the China shock literature, did not seem fully satisfactory; there are robots and Chinese imports in Britain, France, and Germany, too.

A more promising account lies in the fact that America's safety net is so much weaker than the value-added tax–supported welfare systems in Europe. The United States has long been less dependent on trade than smaller European countries and so has had less need to develop a safety net to shield its citizens against trade-related shocks (Rodrik 1998). The baleful effects of racial animus have also undermined attempts to construct a welfare state, so that the United States is left uniquely unprotected in the age of hyper-globalization. We suspect that this is part of the story.

The other thing that differentiates the United States from other rich countries is its health care system, both in terms of its exceptional cost—according to the Organisation for Economic Co-operation and Development (OECD), its share of GDP is about half again as much than the second-most expensive country, Switzerland—and by the way in which it is financed, with most working people covered by insurance provided by their employer.[2]

Case and I know a retired corporate executive who read a draft of our book. His politics are very different from ours, so he brought a perspective that we would not get from most academic audiences; he also knows a great deal about labor practices within large firms. He told us that at one annual meeting with the human resources staff in his firm the executives were told that their health insurance premiums for the next year would increase by 40 percent. Spending that amount was infeasible, so they sent for a well-known firm of management consultants who told them to cut their head count by shedding employees with the highest ratio of health insurance premium to contribution to the firm. Because health insurance premiums do not vary much across employees, that meant getting rid of

2. OECD, "Health Expenditure and Financing," https://stats.oecd.org/Index.aspx?DataSetCode=SHA.

the lowest-paid workers and outsourcing their jobs either domestically or internationally. This is the health wedge in action. It was our friend's belief that the corporation's ex-employees were replaced by illegal immigrants, but his only evidence for this was that they spoke Spanish. Right or not, he was angry, and we suspect that the health wedge has political as well as economic consequences.

The outsourcing is clearly widespread, and few large corporations in the United States employ their own security, transport workers, food service workers, cleaners, or building maintenance staff (Dorn, Schmieder, and Spletzer 2018). These workers, now employed by contract service firms, may be doing work that is close to identical to the work that had been done by employees within the company, but they are paid less, they are less likely to have benefits, and what had been "good jobs" became "not so good jobs." The institutional affiliation, which was important to at least some, has been lost. The new workers may be doing exactly the same work as the original workers, and working alongside the same company employees, but they are not part of the company, and, to use Nick Bloom's words, are not invited to the holiday party (Bloom 2017). This outsourcing of less-skilled workers is one consequence of the health wedge. Another is switching less-skilled workers from full-time to part-time, for whom health insurance need not be offered. To take a local example, the public library in Princeton, New Jersey, not exactly a poor town, complains that many of its (once full-time) staff are now part-timers, and health premiums are cutting into its budget for books.

Finkelstein, McQuillan, Zidar, and Zwick provide a much fuller analysis than we did. We did not think of the effects on the college wage premium, nor did we calculate the differential employment effects. I am delighted that they have done so, but mostly I am pleased and relieved that work on the health wedge exists and is ongoing. Before our book was published, and immediately afterward, we tried hard to sell the idea to friendly labor economists. We pressed hard on the contrast between the vast literature on the minimum wage and the absence of literature on the health wedge. The magnitudes are similar, as is confirmed by the authors. The average annual health insurance premium for a single person cost $7,911 in 2022 and varies little across different employees (Claxton and others 2022). That is $4.00 an hour for a 2,000-hour year; for a worker on the federal minimum wage, health insurance adds 55 percent to labor costs. For the remuneration of executives in the C-suite, the cost is negligible. Yet we were met by blank stares and incomprehension. It was only when Saez and Zucman (2019) published *The Triumph of Injustice*, after our book had gone to press, that

we discovered we were not alone in the health wedge universe. Finkelstein, McQuillan, Zidar, and Zwick take that process several important steps further.

Among the most important findings of their calibration study is that the health wedge is quantitatively important. Compared with a tax proportional to the wage bill, and with plausible functional forms and parameter settings, it reduces employment of people without a college degree by 500,000. It also reduces their earnings by 3 percent so that the college wage premium is increased by 11 percent. Somewhat larger but similar effects come from reducing the cost of the system to Canadian levels. Also important is their finding that the likely size of these effects is comparable to those of other forces that have been arguably harming the well-being of the American working class, including globalization, robots, outsourcing, the decline of unions, and the fall in the real value of the minimum wage. Notably, their estimated magnitudes are not very sensitive to variations in their assumptions. Of course, less-educated workers themselves are less concerned about which of these catastrophes is more or less important; they must suffer all of them. I will return to this point below.

The authors do not consider it, but we might go further and look at a financing scheme based on value-added, including profits as well as earnings, which we can think of as income tax financing. A payroll tax exempts capital from contributing to the cost of health care, so that, following along the same analysis as in the paper, there is also a wedge that favors capital over both kinds of labor. Compared with an income tax, current flat-tax financing creates inequality, not just between more- and less-skilled workers, but also between workers and the owners of capital.

Employer-provided health insurance raises another inequality-related issue, this time on measurement. Because the employer-provided insurance is valuable to employees, what the authors refer to as its amenity value, there is a case for including that amenity value in earnings. Standard measures of wage inequality do not do so, but the Congressional Budget Office (CBO), in its reports on income inequality, adds employer-provided health benefits to household income.[3] Those of us who consider the cost of health care to be much larger than necessary regard part of these payments as extortion by the industry and would argue for their inclusion in industry profits, not in household income. The CBO's accounting thus understates the degree of household inequality. The excessive size of these payments is hurting workers so that to add them to household income is adding statistical insult to actual injury.

3. For the 2019 version, see Congressional Budget Office (2022).

That employer-financed health care works like a head tax, and that this fact is essentially unrecognized by policymakers and economists alike, points again to the long-recognized lack of transparency in the way American health care is financed. Many employees do not recognize that they are paying anything for their insurance and regard it as a gift from their employers. This shrouding makes it very difficult to have rational discussions of reform, something that is very much in the interests of those who are benefiting from health care's extraordinarily high costs. It is surely inconceivable in an open, well-informed, and democratic discussion of health care that a head tax on labor would ever be put in place. It is not the least of the contributions of this paper that it penetrates some of the fog.

It would be much easier to reform health care financing if the system were useless, but it is not. Many years ago, Cutler (2008) documented what happened after President Nixon's 1971 declaration of war on cancer. Mortality from cancer increased by 8 percent over the next two decades, twice the rate of increase over the previous two decades. But since 1990, cancer mortality has fallen across most types of cancer, and although reductions in cigarette smoking have been important, there is also no doubt that the medical system has been important too, with new and effective drugs and with increasingly prevalent screening. Mortality from cancer has fallen since 1992 for both men and women, for those both with and without a college degree, though by more for the former than for the latter (Case and Deaton 2023).

We do not want to threaten this continuing progress, and there is likely more to come, for example, through treatments for Alzheimer's disease or for obesity. Apart from guaranteeing access to treatment, health insurance is important for peace of mind and to protect our pockets against the costs of expensive episodes of sickness. Calculations such as those by Dzau and others (2017) or by Shrank, Rogstad, and Parekh (2019) suggest that America's unnecessary expenditures on health care are on the order of a trillion dollars. Similar numbers come from imagining what would be the case if the United States were to spend not the 19 percent of GDP it spent in 2020 but the 12 percent of GDP that was spent in Switzerland, whose mortality outcomes are much better than ours. These savings would be enough to finance either America's military or America's education system, in either case with change left over. The prices of health care goods and services in the United States are multiples of those in other countries, and the American system is a heavy provider of low-value but highly profitable procedures. As always, resources lost to waste and abuse are not simply shot into space but show up as someone's salary (or stock portfolio). The

system is effective at redistributing wealth upward from firms and their employees, especially less-skilled employees, to doctors, device manufacturers, and pharmaceutical company and hospital executives.

The burden of health care falls on us all, if not equally. It is not costless to waste a trillion dollars a year. Not only are wages lower than they need to be but so are profits. In 2013, long before COVID-19 added to the deficit, Blinder (2013, 405) wrote that "if we can somehow solve the health care cost problem, we will also solve the long-run deficit problem. But if we can't control health care costs, the long-run deficit problem is insoluble." All of this has dire effects on politics, not just on the economy.

I want to direct the remainder of my remarks to measures of the well-being of Americans without a college degree. As the authors show, the health wedge means that the burden of health care financing differentially affects the noncollege-educated, but this is not the only force that is working against them. They note that the effects of the wedge are comparable in size to those from the decline in private-sector unions, from trade, from automation, and from declines in the federal minimum wage. We cannot simply add up these effects, but I do not need to sort out the ways they overlap to know that the cumulative effect will be much larger than the effects of the wedge that are discussed in the paper. Figure 2 in the paper addresses outcomes for college-educated and noncollege-educated workers, and I want to focus on two of the measures shown, real earnings, in panel A, and the employment rate, in panel C.

Starting with panel C, on employment, this shows men and women together, and because women's participation in the labor force was rising until around 2000 and falling thereafter, we get a different picture if we look at men and women separately.[4] For noncollege-educated men, attachment to the labor force shows a long-term downward trend punctuated by upturns when the economy is doing well, but the peak to which each upturn leads is always lower than the previous one. Even in the best of times, participation of less-educated men is lower than it was in the previous best of times. (It is possible that this pattern has been broken post-COVID-19, but it is too early to tell.) I believe that this long-term decline comes not from any decline in supply associated with a loss of virtue (Murray 2012) but from the cumulative loss of good jobs for working-class men, a loss that is

4. Figure 11.2 in Case and Deaton (2020) shows the employment rates for 25- to 54-year-olds by college degree status and gender; the profile for women without a college degree follows an inverted *U*-shape, rising until 2000, so that when women and men are taken together, the rise in women's participation annihilates the fall for men up to 2000.

exacerbated by the rising cost of health care worsened by the health wedge. This is part of the general loss of institutional attachment among working-class men documented by Edin and others (2019).

Measuring employment is relatively straightforward compared with measuring real wages. Incomes, expenditures, and wages are more complicated to assess than a binary indicator of whether someone is employed. Beyond the estimation of money amounts, which is often difficult in itself, we need price indexes to correct for inflation.

Panel A of figure 2 in the paper shows estimates of real earnings for those with and without a college degree. There is a substantial divergence with more rapid growth for the more-educated, which is what drives the expansion of the college wage premium in panel B. As the accompanying notes point out, the increase of real wages of $7,754 from 1977 to 2017 becomes a decrease in real wages of $1,779 if, instead of using the Personal Consumption Expenditures (PCE) price index to convert nominal to real, we use the Consumer Price Index (CPI). Over the forty years from 1980 to 2020, the CPI rose at 2.9 percent a year and the PCE index by 2.41 percent per year; over forty years, there was 16 percent more inflation in the CPI. Over four decades, we might argue that a difference of 16 percent is not of great importance. But that depends on the purpose to which it is being put. It is certainly of rhetorical significance. The statement "real wages of workers without a bachelor's degree are lower now than forty years ago" is true when the CPI is used to deflate and false if the PCE price index is used to deflate. Another important measure that depends on the CPI is the poverty rate, and its behavior over time is markedly different using the PCE price index (Jencks 2015). Social Security payments would also grow less rapidly if the PCE price index were to substitute for the CPI.

There is no straightforward answer to which index is right and, as I shall argue, even if the choice were settled, more difficult issues lurk beyond. The Bureau of Economic Analysis publishes a table that shows step-by-step reconciliation between the two indexes, and I draw on that and on other work here.[5]

The PCE price index is a Fisher index while the CPI is a Laspeyres index, and many economists prefer the former to the latter. If the two indexes used identical prices and quantities—which they do not—the Fisher would respond to the negative correlation between changes in prices

5. Bureau of Economic Analysis, "National Income and Product Accounts," table 9.1U: Reconciliation of Percent Change in the CPI with Percent Change in the PCE Price Index, https://apps.bea.gov/iTable/?reqid=19&step=3&isuri=1&1921=underlying&1903=2075.

and changes in quantities by rising less rapidly than the Laspeyres. Arguably, this makes the Fisher a better approximation of a true cost-of-living index, but that welfare result is not readily established, not least because it would apply to a representative agent, not to any specific group of consumers, and because there are multiple cost-of-living indexes.[6] But the main reason that the indexes diverge is not to do with formula differences but comes from using different weights (weighting) and from covering different bundles of goods (scope). In effect, the two indexes are measuring different things, the underlying bundles that are being priced are not the same.

The PCE price index, which is a byproduct of the National Accounts, uses expenditures supplied by business. The CPI is produced by the Bureau of Labor Statistics and uses (mainly) household survey data to construct weights, data that come from the Consumer Expenditure Survey. In most countries around the world, and the United States is no exception, expenditures from household surveys are far from a perfect match to expenditures in the National Accounts, and although progress has been made to reconcile them, differences remain (Passero, Garner, and McCully 2015). Even were the surveys to match the National Accounts, differences in scope remain. Perhaps the most important of these is that the CPI weight for health care covers only expenditures made by consumers and excludes expenditures made on behalf of consumers by nonprofits, by employers, and by government. In consequence, the PCE price index weight for health care is larger than the CPI weight for health care, which implies that the next most important item, housing, has a larger weight in the CPI. An important source of difference between the two indexes is thus the extent to which the two prices are rising at different rates. Johnson (2017) shows that between 2014 and 2016, the PCE price index increased at 1.25 percent annually, compared with only 1.06 percent for the CPI, largely because of the more rapid increase in the price of health care. He notes that indexing Social Security on the PCE price index instead of the CPI would have led to $10 billion additional government expenditure over the period. For an earlier period, from 2002 to 2007, McCully, Moyer, and Stewart (2007) show that the weight and scope effects approximately cancel, with most of the faster inflation in the CPI accounted for by the formula effect.

Clearly, the hunt for the right price index is fraught with difficulty, which raises the specter of unprincipled political interference in what many might hope would be a technical matter. Politicians stand ready to step into the gaps left by statisticians.

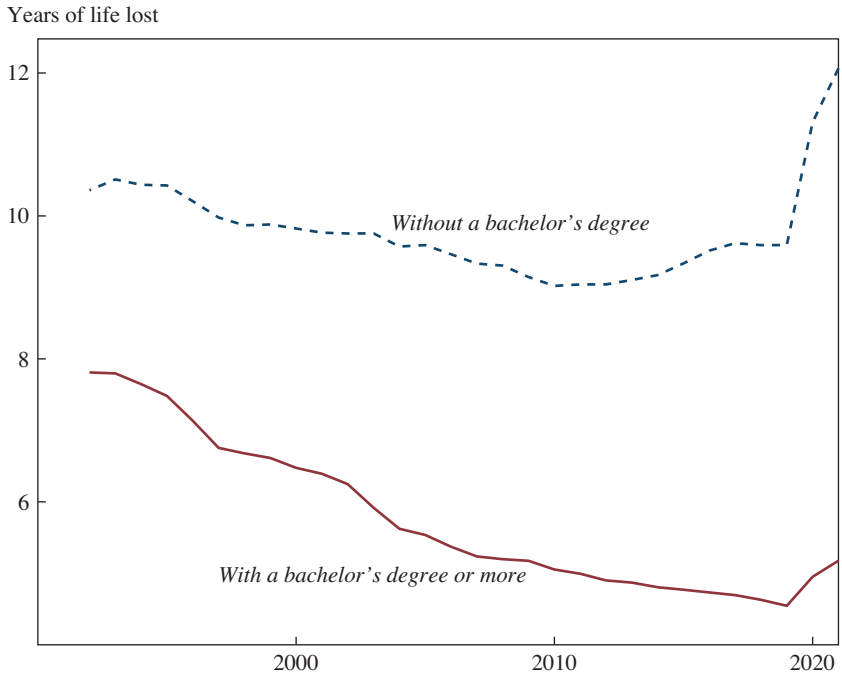
6. See, for example, Schultze and Mackie (2002).

I have not yet mentioned perhaps the most difficult issue besetting price indexes, which is how to treat quality change. In some cases, quality changes are isomorphic to quantity changes. If an improved gasoline were to deliver twice as many miles per gallon as the original, it effectively costs half as much, and the obvious (and obviously correct) procedure is to enter into the index not its price but half of its price. This quantity-augmenting analogy serves as the basis for quality correction in both the PCE price index and the CPI, even though its validity is not always clear. The original good is not always available, so consumers may be forced to buy a more expensive good or service than they need, so that for them quality correction understates the true increase in price. Cass (2023) has calculated what he calls a "cost-of-thriving index" by calculating how many hours of median-wage work is needed to buy a list of important goods for middle-class Americans, including a house, a car, and health care; his index rises much more rapidly than either the CPI or the PCE price index. He refers to "the catastrophic erosion of middle-class life in America" (Cass 2023, 6). Cass's index is unlikely to replace standard indexes, if only because it covers only some items in the budget, but his work is thought-provoking—it highlights the conceptual difficulties of statistical agencies' current hedonic correction procedures and illustrates (if likely overstating) the dangers of quality correction.

The argument that both the PCE price index and the CPI overstate the true increases in prices was made in an influential Advisory Commission report, commissioned by the Senate Finance Committee and commonly referred to, after its chairman, as the Boskin Report (Advisory Committee to Study the Consumer Price Index 1996). The report argued that the CPI was overstating the rate of inflation by 1.5 percentage points a year, of which 1.0 percent came from unmeasured quality change, and recommended that, going forward, CPI inflation be reduced by this amount each year. The report makes a brave (or perhaps foolhardy, depending on the reader) attempt to estimate quality change for each category of consumer expenditure, often when there was no underlying research to support the estimate. The Bureau of Labor Statistics responded effectively, noting that they make adjustments when there is good evidence but that national statistical offices cannot rely on guesswork. A National Academy report (Schultze and Mackie 2002) came to a similar conclusion, but a number of writers continue to make the correction called for in the Boskin Report, particularly in nonofficial measures of poverty.[7]

7. See Deaton (2023, chapters 3 and 4) for an informal account.

Figure 1. Expected Years of Life Lost between Age 25 and 85



Source: Data from Case and Deaton (2023).

Perhaps it is impossible to estimate fully credible estimates of inflation, or at least to do so in a way that commands close to universal assent, let alone assent across the political divide. Which is why, in our own work, Case and I have turned to measures of well-being that are not subject to these kinds of problems. In particular, whether someone is alive or dead is typically clear and (mostly) unchallengeable. Even if there is wide disagreement on whether the American working-class is prospering or suffering, evidence that they are dying is hard to explain away, especially when they are dying by suicide or by drug or alcohol overdose. Clearly, something is badly wrong.

I therefore conclude with some of the evidence on mortality. Figure 1 shows one of many possible measures: the expected number of years of life lost (YLL) between an individual's twenty-fifth and eighty-fifth birthdays. The measure is calculated following standard life table principles. In each year, we have a set of age-specific mortality rates for people with and without a college degree, and we collapse those into a single number,

the expected years of life from age 25 to 85 at that year's mortality rates. Expected YLL is 60, the maximum, minus the expected years of life lived. Note that this does not assume that the mortality rates hold constant going forward; it is simply a way of aggregating the sixty age-specific mortality rates from each year into a single number. A similar procedure, with similar results, would be to calculate age-adjusted mortality, though this requires a set of population weights, which is not required to calculate YLL. The underlying data are those used in Case and Deaton (2023), which explains our procedures in more detail. The figure covers men and women together, and all racial and ethnic groups.

The bottom line, for those with a four-year college degree, shows continuous decline from 1992 to 2019, with the rate of decline slowing somewhat after 2010, followed by a marked upturn during the pandemic. The top line, for those without a bachelor's degree, shows mortality decline up to 2010, albeit at a slower rate than for those with a bachelor's degree, but the mortality decline reverses after 2010 so that, even before the pandemic, the mortality experiences of those with and without a degree were going in opposite directions. During the pandemic, the less-educated lost two and a half years of expected life, compared with 0.2 years for those with a degree. The YLL gap between the more- and less-educated group doubled between 1992 and 2019, from 2.5 years to 5 years, and then jumped to 6.9 years in 2021.

Figure 1 here echoes figure 2 in the paper, though it matches most closely a version of panel C in which real wages of the noncollege-educated are falling; of course, we are not claiming that mortality is an alternative measure of real wages or that real wages are indeed falling.

The many details and a few caveats to figure 1 are covered in Case and Deaton (2023), though that paper does not work with YLL. And since it is often asked, it is true that the fraction of Americans with a bachelor's degree rose between 1992 and 2021, so that there was likely some health-based selection from the less- to the more-educated group over time. Such selection increases mortality for both groups and can either widen or narrow the gap. The evidence suggests that the gap is little affected by health selection into education.

REFERENCES FOR THE DEATON COMMENT

Advisory Commission to Study the Consumer Price Index. 1996. *Toward a More Accurate Measure of the Cost of Living: Final Report to the Senate Finance Committee*. Washington: Social Security Administration. https://www.ssa.gov/history/reports/boskinrpt.html.

Blinder, Alan S. 2013. *After the Music Stopped: The Financial Crisis, the Response, and the Work Ahead*. New York: Penguin Books.

Bloom, Nicholas. 2017. "Corporations in the Age of Inequality." *Harvard Business Review*, March 21. https://hbr.org/cover-story/2017/03/corporations-in-the-age-of-inequality.

Case, Anne. 2015. "'Deaths of Despair' Are Killing America's White Working Class." *Quartz*, December 30. https://qz.com/583595/deaths-of-despair-are-killing-americas-white-working-class.

Case, Anne, and Angus Deaton. 2015. "Rising Mortality and Morbidity in Midlife among White Non-Hispanic Americans in the 21st Century." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 49: 15078–83.

Case, Anne, and Angus Deaton. 2020. *Deaths of Despair and the Future of Capitalism*. Princeton, N.J.: Princeton University Press.

Case, Anne, and Angus Deaton. 2022. "The Great Divide: Education, Despair and Death." *Annual Review of Economics* 14:1–21.

Case, Anne, and Angus Deaton. 2023. "Accounting for the Widening Mortality Gap between American Adults with and without a BA." *Brookings Papers on Economic Activity*, Fall, conference draft. https://www.brookings.edu/wp-content/uploads/2023/09/1_Case-Deaton_unembargoed.pdf.

Cass, Oren. 2023. "The 2023 Cost-of-Thriving Index: Tracking the Catastrophic Erosion of Middle-Class Life in America." Washington: American Compass. https://americancompass.org/wp-content/uploads/2023/02/COTI_2023_Final.pdf.

Claxton, Gary, Matthew Rae, Emma Wager, Gregory Young, Heidi Whitmore, Jason Kerns, Greg Shmavonian, and Anthony Damico. 2022. *2022 Employer Health Benefits Survey*. San Francisco: Kaiser Family Foundation.

Congressional Budget Office. 2022. *The Distribution of Household Income, 2019*. Washington: Author. www.cbo.gov/publication/58353.

Cutler, David M. 2008. "Are We Finally Winning the War on Cancer?" *Journal of Economic Perspectives* 22, no. 4: 3–26.

Deaton, Angus. 2023. *Economics in America: An Immigrant Economist Explores the Land of Inequality*. Princeton, N.J.: Princeton University Press.

Dorn, David, Johannes F. Schmieder, and James R. Spletzer. 2018. "Domestic Outsourcing in the United States." https://www.dol.gov/sites/dolgov/files/OASP/legacy/files/Domestic-Outsourcing-in-the-United-States.pdf.

Dzau, Victor J., Mark B. McClellan, J. Michael McGinnis, Sheila P. Burke, Molly J. Coye, Angela Diaz, and others. 2017. "Vital Directions for Health and Health Care: Priorities from a National Academy of Medicine Initiative." *JAMA* 317, no. 14: 1461–70.

Edin, Kathryn, Timothy Nelson, Andrew Cherlin, and Robert Francis. 2019. "The Tenuous Attachments of Working-Class Men." *Journal of Economic Perspectives* 33, no. 2: 211–28.

Jencks, Christopher. 2015. "The War on Poverty: Was It Lost?" *New York Review*, April 2. https://www.nybooks.com/articles/2015/04/02/war-poverty-was-it-lost/.

Johnson, Noah. 2017. "A Comparison of PCE and CPI: Methodological Differences in U.S. Inflation Calculation and Their Implications." Washington: Bureau of Labor Statistics. https://www.bls.gov/osmr/research-papers/2017/pdf/st170010.pdf.

McCully, Clinton P., Brian C. Moyer, and Kenneth J. Stewart. 2007. "A Reconciliation between the Consumer Price Index and the Personal Consumption Expenditures Price Index." Washington: Bureau of Economic Analysis and Bureau of Labor Statistics. https://www.bea.gov/system/files/papers/P2007-4.pdf.

Murray, Charles. 2012. *Coming Apart: The State of White America, 1960–2010*. New York: Crown Publishing Group.

Passero, William, Thesia I. Garner, and Clinton P. McCully. 2015. "Understanding the Relationship: CE Survey and PCE." In *Improving the Measurement of Consumers' Expenditures*, edited by Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus. Chicago: Chicago University Press.

Rodrik, Dani. 1998. "Why Do More Open Economies Have Bigger Governments?" *Journal of Political Economy* 106, no. 5: 997–1032.

Saez, Emanuel, and Gabriel Zucman. 2019. *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay.* New York: W. W. Norton.

Schultze, Charles L., and Christopher Mackie, eds. 2002. *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*. Washington: National Academy Press.

Shrank, William H., Teresa L. Rogstad, and Natasha Parekh. 2019. "Waste in the US Health Care System: Estimated Costs and Potential for Savings." *JAMA* 322, no. 15: 1501–9.

GENERAL DISCUSSION Matthew Fiedler offered a critique of the model structure. Fiedler noted that the authors provide a representative firm that offers health care coverage, which contrasts with reality. In the real world, outside of the model, Fiedler claims, there are many firms that do not offer coverage and these firms disproportionately employ noncollege workers. As a result, Fiedler expected the wages of the noncollege-educated workers to be pinned down by the non-offering firms, which may seriously constrain the ability of the offering firms to pass through the costs to their noncollege-educated workers. Thus, Fiedler suspected the real world may not be well represented by the head tax financing modeled by the authors.

Adam Looney further expressed skepticism of the characterization of the health care wedge as a fixed per worker cost whose rising cost thereby consumed a larger share of lower-paid workers' compensation for three reasons. First, he noted Brooks Pierce's paper "Compensation Inequality" and subsequent work suggests that compensation inequality is increasing at the same pace as wage inequality, which suggests that the health care cost

wedge is rising faster for higher-paid workers.[1] Second, health insurance is in reality quite progressive rather than being a flat tax. The Affordable Care Act (ACA), for example, requires all employers to provide the values of the premiums that they pay to their W-2 employees. Analysis of these data shows that the top 10 percent of employees account for 21 percent of health insurance premiums.[2] Thus, while the existing "tax" of insurance premiums may not be as progressive as a payroll tax would be, it is closer to a payroll tax than a poll tax. Lastly, looking over time, one important consequence of rising health insurance costs was that it encouraged employers to drop coverage, particularly among lower-paid workers. While that may have reduced their total compensation, it also meant it was less of a "wedge" reducing their potential wages or employment prospects.

Eswar Prasad suggested broadening the study to include other fixed or quasi-fixed costs in the benefit package, such as Social Security contributions. Prasad pointed out that including these other fixed or quasi-fixed costs would amplify the findings of this study. Further, he agreed with the point made by Looney that compensation and wage inequality have tracked together in overall trend. Unlike Looney, however, Prasad suggested that the nonwage benefits may alter the dynamics of the model presented by the authors. John Haltiwanger, in a later remark, agreed that there are potentially multiple levers at work that are not necessarily mutually exclusive and suggested that the authors consider interactions between the various levers.

Haltiwanger expressed simultaneous skepticism and interest in whether the authors account for empirical patterns. He primarily focused on advocating for using Bureau of the Census data rather than Social Security Administration (SSA) data, in part because the SSA data used by Song and others experience a high degree of missingness for key variables.[3] Haltiwanger, in a paper with Henry Hyatt and James Spletzer, used the industry codes and found that most of the rising inequality between firms was actually rising inequality between industries.[4] But even more striking,

1. Brooks Pierce, "Compensation Inequality," *Quarterly Journal of Economics* 116, no. 4 (2001): 1493–525; and Brooks Pierce, "Recent Trends in Compensation Inequality," in *Labor in the New Economy*, Katharine G. Abraham, James R. Spletzer, and Michael J. Harper, eds. (Chicago: University of Chicago Press, 2010).

2. Ithai Z. Lurie and Corbin L. Miller, "Employer-Sponsored Health Insurance Premiums and Income in U.S. Tax Data," *Journal of Public Economics* 224 (2023): 104942.

3. Jae Song, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter, "Firming Up Inequality," *Quarterly Journal of Economics* 134, no. 1 (2019): 1–50.

4. John C. Haltiwanger, Henry R. Hyatt, and James R. Spletzer, "Industries, Mega Firms, and Increasing Inequality," working paper 29920 (Cambridge, Mass.: National Bureau of Economic Research, 2022).

only 10 percent of the approximately three hundred four-digit industries were responsible for the bulk of this trend. Of the thirty industries that might be considered "household name" industries, there are nineteen at the top that are high-tech industries, finance, and physicians. At the bottom, there are eleven industries, including general merchandise stores, grocery stores, and restaurants. Regarding David Cutler's comment on the decline in earnings at megafirms, that decline is observed at the eleven bottom industries. So it's plausible that the rising cost of health care and the authors' hypothesis is limited to a very distinct pattern in a very small number of industries.

Neil Mehrotra and Melissa Kearney both discussed the college wage premium. Mehrotra asked a specific mechanical question regarding the dynamics of the college wage premium: To the extent that the authors see US health care expenditures relative to other countries as a subsidy to providers, would that mechanically increase the college wage premium, to the extent that most providers are college-educated? Kearney presented comments suggesting that the authors supplement their theoretical calibration with an empirical study. Kearney anticipated that post-2010, after changes implemented by the ACA and the expansion of Medicaid, in equilibrium we should have seen fewer low-wage, noncollege-educated workers getting health insurance through their employers, which would have led to a shrinking of the college wage premium.

Adele Morris made positive comments placing the authors' work in the broader context of inefficient and negative outcomes in other policy arenas. These include climate change policy's unpopularity due to the lack of a well-designed safety net that would mitigate the impacts of contractionary pro-climate legislation. In this respect, Morris called the United States a political outlier.

The authors provided responses in conclusion. Owen Zidar first addressed Angus Deaton's comments on the value of insurance. He noted that the authors do not need the value of insurance for the payroll tax counterfactual, rather only for the Canada counterfactual. Zidar further addressed Kearney's comment, noting that the original idea for the paper involved more estimation and empirical components. The authors initially planned to use regional variation in health prices to examine places that, due to consolidation or innovation, had higher health care costs and to investigate the amenity value of these higher costs. Zidar further noted that comparing technological improvements driving cost versus increasing administrative costs would be a challenge beyond the scope of the paper. Zidar further noted, as Cutler and Deaton suggested, that if the value of health insurance is low, then there is less amenity effect, and it is mostly observed as a tax increase and

the distortion grows with the square of the tax rate. Thus, as health care costs go up, the distortion will increase. Finally, Zidar agreed with Deaton's critique of the correct inflation metric to use for the study, and the authors will make corrections regarding the CPI.

Amy Finkelstein noted that the general theme of the comments was "this can't be all that is going on" because the phenomenon described by the authors doesn't match an exact time series or the between-industry facts. The authors agree with this criticism, as the goal of this exercise was to evaluate whether the health care wedge was quantitatively important and to determine whether it should be studied more. Finkelstein cited books by Case and Deaton and by Saez and Zucman from the introduction of the paper as ample sources of qualitative evidence that this project was a good line of inquiry.[5] Finkelstein further referenced the provisional results in the form of supply-demand graphs Zidar presented to provide evidence that the question may be quantitatively important. Provisionally, the authors' answer is that the health care wedge could indeed be quantitatively important, and thus invites further study. Additionally, Finkelstein emphasized the importance of health care costs in labor inequality in general and provided a counterargument to Kearney's question. She noted the authors attempted empirical work on this question, and they did attempt to investigate the Medicaid expansion. However, they found that the Medicaid expansion primarily took people out of the ranks of the previously uninsured, surprisingly to some, rather than those already insured by their employer. In terms of other types of benefits, as mentioned by Prasad, Finkelstein acknowledged their importance but noted that Social Security is already financed as a payroll tax.[6] Thus, there is not the same lump sum per worker phenomenon for Social Security as Prasad suggested.

Eric Zwick concluded the authors' remarks, noting that the ACA, when introduced, had provisions attempting to prevent firms from outsourcing but this did not reach to international outsourcing.[7] Thus, as referenced

5. Anne Case and Angus Deaton, *Deaths of Despair and the Future of Capitalism* (Princeton, N.J.: Princeton University Press, 2020); and Emmanuel Saez and Gabriel Zucman, *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay* (New York: W. W. Norton, 2019).

6. Social Security Administration, "How Is Social Security Financed?," https://www.ssa.gov/news/press/factsheets/HowAreSocialSecurity.htm#:~:text=Social%20Security%20is%20financed%20through,self%2Demployed%20pay%2012.4%20percent.&text=The%20payroll%20tax%20rates%20are,up%20to%20a%20certain%20amount.

7. Christopher G. Guldberg and Michael J. Poland, "The Affordable Care Act: International Implications," *Lexology*, March 18, 2013, https://www.lexology.com/library/detail.aspx?g=c87ac028-18d8-4f78-984f-930a1f03770a.

by Haltiwanger, some of the industry-specific trends and types of workers exposed to outsourcing over a twenty-to-thirty-year period will be middle income, older workers in higher health cost industries such as heavy manufacturing. Zwick noted that this is more representative of the quantities rather than earnings piece. For some of the other response margins, however, quantities of employment were more relevant. Thus, Zwick noted, the earnings data are partially incomplete, and thus, regarding compensation, there may be an overstating of welfare inequality. Zwick further addressed Looney's point, noting there are two offsets to take into account and referring back to Pierce's paper.[8] Zwick noted that if the authors had focused on health care alone, compensation inequality would have gone up less than pay inequality because the health piece is much larger in the middle of the distribution, with the offsets being pensions and 401(k)s in the top portion of the earning distribution, which Zwick determined was beyond the scope of the paper.

8. Brooks Pierce, "Recent Trends in Compensation Inequality," in *Labor in the New Economy*, Katharine G. Abraham, James R. Spletzer, and Michael J. Harper, eds. (Chicago: University of Chicago Press, 2010).