Janice Eberly and Jón Steinsson, Editors

Brookings Papers

ON ECONOMIC ACTIVITY

SPRING 2024

STANTCHEVA
Why Do We Dislike Inflation?

ATKESON and KISSLER
The Impact of Vaccines and Behavior on US Cumulative Deaths from COVID-19

ARSLANALP, EICHENGREEN, and HENRY
Sustained Debt Reduction: The Jamaica Exception

AFROUZI, HALAC, ROGOFF, and YARED
Changing Central Bank Pressures and Inflation

FIELDHOUSE, MUNRO, KOCH, and HOWARD
The Emergence of a Uniform Business Cycle in the United States: Evidence from New Claims-Based Unemployment Data

HANSON, IVASHINA, NICOLAE, STEIN, SUNDERAM, and TARULLO
The Evolution of Banking in the 21st Century: Evidence and Regulatory Implications

Brookings Papers ON ECONOMIC ACTIVITY

BPEA
SPRING
2024

BROOKINGS

PUBLISHED FOR THE BROOKINGS INSTITUTION BY
JOHNS HOPKINS
UNIVERSITY PRESS

Brookings Papers

ON ECONOMIC ACTIVITY

SPRING 2024

JANICE EBERLY
JÓN STEINSSON
Editors

PUBLISHED FOR THE BROOKINGS INSTITUTION BY
JOHNS HOPKINS
UNIVERSITY PRESS

Brookings Papers on Economic Activity
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.

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

Brookings Papers

ON ECONOMIC ACTIVITY

SPRING 2024

STEFANIE STANTCHEVA
Why Do We Dislike Inflation? 1
Comment by Carola Binder 47
Comment by Yuriy Gorodnichenko 54
General Discussion 60

ANDREW ATKESON and STEPHEN KISSLER
The Impact of Vaccines and Behavior on US Cumulative Deaths from COVID-19 67
Comment by Thomas Philippon 113
Comment by Coady Wing 116
General Discussion 128

SERKAN ARSLANALP, BARRY EICHENGREEN, and PETER BLAIR HENRY
Sustained Debt Reduction: The Jamaica Exception 133
Comment by Laura Alfaro 182
Comment by Emil Verner 192
General Discussion 202

HASSAN AFROUZI, MARINA HALAC, KENNETH ROGOFF, and PIERRE YARED
Changing Central Bank Pressures and Inflation 205
Comment by Donald Kohn 242
Comment by Silvana Tenreyro 249
General Discussion 261

ANDREW J. FIELDHOUSE, DAVID MUNRO, CHRISTOFFER KOCH, and SEAN HOWARD
The Emergence of a Uniform Business Cycle in the United States: Evidence from New Claims-Based Unemployment Data 265
Comment by Steven J. Davis 320
Comment by Marianna Kudlyak 326
General Discussion 337

SAMUEL G. HANSON, VICTORIA IVASHINA, LAURA NICOLAE,
JEREMY C. STEIN, ADI SUNDERAM, and DANIEL K. TARULLO
*The Evolution of Banking in the 21st Century:
Evidence and Regulatory Implications* 343
Comment by Arvind Krishnamurthy 390
Comment by Raghuram Rajan 396
General Discussion 406

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; Smith Richardson Foundation; and State Farm Mutual Automobile Insurance Company. We express appreciation to Brevan Howard Research 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. We thank Paris Rich Bingham, Janina Bröker, Alexander Conner, Olivia Howard, and Liam Marshall for preparing the summaries of general discussions for this volume.

 Conference drafts and recordings of the Spring 2024 *BPEA* Conference can be accessed at https://www.brookings.edu/events/bpea-spring-2024-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 https://www.brookings.edu/projects/brookings-papers-on-economic-activity/all-bpea-research/. 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 EIGHTEENTH CONFERENCE

Henry Aaron *Brookings Institution*
Katharine Abraham *University of Maryland*
Hassan Afrouzi *Columbia University*
George Akerlof *Georgetown University*
Laura Alfaro *Harvard Business School*
Serkan Arslanalp *International Monetary Fund*
Andrew Atkeson *University of California, Los Angeles*
Alan Auerbach *University of California, Berkeley*
Laurence Ball *Johns Hopkins University*
Jose Maria Barrero *Instituto Tecnológico Autónomo de México*
Susanto Basu *Boston College*
Joe Beaulieu *Brevan Howard*
Ben Bernanke *Brookings Institution*
Carola Binder *University of Texas at Austin*
Hoyt Bleakley *University of Michigan, Ann Arbor*
Alan Blinder *Princeton University*
Barry Bosworth *Brookings Institution*
William Brainard *Yale University*
Steven Braun *Council of Economic Advisers*
Elaine Buckberg *Harvard Salata Institute for Climate and Sustainability*
Gerald Cohen *UNC Kenan Institute of Private Enterprise*
Susan Collins *Federal Reserve Bank of Boston*
Nicolas Crouzet *Northwestern University Kellogg School of Management*
Steven Davis *Stanford University Hoover Institution*
Ryan Decker *Federal Reserve Board*
Brad DeLong *University of California, Berkeley*
John Driscoll *Federal Reserve Board*
Arindrajit Dube *University of Massachusetts Amherst*
Burcu Duygan-Bump *US Treasury*
Karen Dynan *Harvard University*
Janice Eberly *MIT and Northwestern University*
Wendy Edelberg *Brookings Institution*
Barry Eichengreen *University of California, Berkeley*
William English *Yale School of Management*
Michael Falkenheim *Congressional Budget Office*
Bruce Fallick *Federal Reserve Bank of Cleveland*
Michael Faulkender *University of Maryland*
Andrew Fieldhouse *Texas A&M University*
Kristin Forbes *MIT Sloan School of Management*
Richard Frank *Brookings Institution*
Benjamin Friedman *Harvard University*

Jason Furman *Harvard University*
Joseph Gagnon *Peterson Institute for International Economics*
William Gale *Brookings Institution*
Robert Gordon *Northwestern University*
Egor Gornostay *Peterson Institute for International Economics*
Yuriy Gorodnichenko *University of California, Berkeley*
Josh Gotbaum *Brookings Institution*
Pierre-Olivier Gourinchas *International Monetary Fund*
Carol Graham *Brookings Institution*
Marina Halac *Yale University*
Robert Hall *Stanford University*
John Haltiwanger *University of Maryland*
Samuel Hanson *Harvard Business School*
Benjamin Harris *Brookings Institution*
Peter Blair Henry *Stanford University Hoover Institution*
Doug Holtz-Eakin *American Action Forum*
Sean Howard *Independent*
Caroline Hoxby *Stanford University*
Victoria Ivashina *Harvard Business School*
David Johnson *National Academies of Sciences, Engineering, and Medicine*
Şebnem Kalemli-Özcan *University of Maryland*
Steven Kamin *American Enterprise Institute*
Michael Kiley *Federal Reserve Board*
Stephen Kissler *University of Colorado, Boulder*
Jeffrey Kling *Congressional Budget Office*
Christoffer Koch *Empirical Research Partners*
Donald Kohn *Brookings Institution*
Arvind Krishnamurthy *Stanford University*
Randall Kroszner *University of Chicago Booth School of Business*
Marianna Kudlyak *Federal Reserve Bank of San Francisco*
Adam Looney *Brookings Institution*
Deborah Lucas *Massachusetts Institute of Technology*
Byron Lutz *Federal Reserve Board*
Greg Mankiw *Harvard University*
Jan Kees Martijn *International Monetary Fund*
Mark Mazur *Independent*
Karel Mertens *Federal Reserve Bank of Dallas*
Laurence Meyer *LH Meyer*
Gian Maria Milesi-Ferretti *Brookings Institution*
Silvia Miranda-Agrippino *Federal Reserve Bank of New York*
Jeffrey Miron *Harvard University*
David Munro *Middlebury College*
Laura Nicolae *Harvard University*
Maurice Obstfeld *Peterson Institute for International Economics*
Athanasios Orphanides *Massachusetts Institute of Technology*
Anna Paulson *Federal Reserve Bank of Chicago*
George Perry *Brookings Institution*
Thomas Philippon *New York University*
Jonathan Pingle *UBS*
Richard Portes *London Business School*
Wendell Primus *Brookings Institution*
Raghuram Rajan *University of Chicago Booth School of Business*
Sarah Reber *Brookings Institution*
Tristan Reed *World Bank*
Ricardo Reis *London School of Economics*
Giovanni Ricco *Ecole Polytechnique*
Kenneth Rogoff *Harvard University*

Christina Romer *University of California, Berkeley*
David Romer *University of California, Berkeley*
John Sabelhaus *Brookings Institution*
Natasha Sarin *Yale University*
Jake Schild *Bureau of Labor Statistics*
Benjamin Schoefer *University of California, Berkeley*
Louise Sheiner *Brookings Institution*
Yongseok Shin *Washington University in St. Louis*
Tara Sinclair *US Treasury*
David Skidmore *Brookings Institution*
Stefanie Stantcheva *Harvard University*
Jeremy Stein *Harvard University*
Mark Steinmeyer *Smith Richardson Foundation*
Jón Steinsson *University of California, Berkeley*
James Stock *Harvard University*
Adi Sunderam *Harvard Business School*
Daniel Tarullo *Harvard Law School*
Silvana Tenreyro *London School of Economics*
Robert Tetlow *Federal Reserve Board*
Filiz Unsal *Organisation for Economic Co-operation and Development*
Eric Van Nostrand *US Treasury*
Emil Verner *Massachusetts Institute of Technology*
Stan Veuger *American Enterprise Institute*
Alan Viard *American Enterprise Institute*
Gianluca Violante *Princeton University*
David Wessel *Brookings Institution*
David Wilcox *Peterson Institute for International Economics*
Coady Wing *Indiana University*
Justin Wolfers *University of Michigan*
Susan Woodward *Independent*
Abigail Wozniak *Federal Reserve Bank of Minneapolis*
Pierre Yared *Columbia University*
Eric Zwick *University of Chicago Booth School of Business*

Paris Rich Bingham *Brookings Institution*
Janina Bröker *Brookings Institution*
Haowen Chen *Brookings Institution*
Alexander Conner *Brookings Institution*
Siobhan Drummond *Brookings Institution*
Olivia Howard *Brookings Institution*
Liam Marshall *Brookings Institution*

STEFANIE STANTCHEVA
Harvard University

Why Do We Dislike Inflation?

ABSTRACT This paper provides new evidence on a long-standing question asked by Shiller (1997): why do we dislike inflation? I conducted two surveys on representative samples of the US population to elicit people's perceptions about the impacts of inflation and their reactions to it. The predominant reason for people's aversion to inflation is the widespread belief that it diminishes their buying power, as neither personal nor general wage increases seem to match the pace of rising prices. As a result, respondents report having to make costly adjustments in their budgets and behaviors, especially among lower-income groups. Inflation also provokes stress, emotional responses, and a sense of inequity, as the wages of high-income individuals are perceived to grow more rapidly amid inflation. Many respondents believe that firms have considerable discretion in setting wages, opting not to raise them in order to boost profits, rather than being compelled by market dynamics. The potential positive associations of inflation, such as with reduced unemployment or enhanced economic activity, are typically not recognized by respondents. Inflation ranks high in priority among various economic and social issues, with respondents blaming the government and businesses for it. I also highlight a substantial polarization in attitudes toward inflation along partisan lines, as well as across income groups.

Over twenty-five years ago, Shiller (1997, 13) wanted to "understand, through public survey methods, why people are so concerned and dismayed by inflation." In a nutshell, he discovered that individuals consider inflation a national concern primarily because it undermines their living standards. They observe prices rising while their wages stagnate,

Conflict of Interest Disclosure: The author did not receive financial support from any firm or person for this paper or from any firm or person with a financial or political interest in this paper. 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 2024: 1–46 © 2024 The Brookings Institution.

attributing this imbalance to employers' "greed." Moreover, respondents associated inflation with economic downturns and political instability, citing certain "unspecified systemic factors" (ibid., 57).

Considering the significant time elapsed since this seminal study, it is important to refresh our understanding of the public's aversion to inflation. The COVID-19 pandemic has thrust inflation back into the limelight as a critical policy issue in the United States and abroad, reigniting concerns over its effects on living standards. Given the transformations our economic system has experienced since the late 1990s, including the impact of globalization, the financial crisis, the pandemic, and a growing polarization in societal perspectives (Alesina, Miano, and Stantcheva 2020), contemporary views on inflation and the economy might have shifted significantly . . . or have they?

Drawing inspiration from Shiller (1997), this paper offers an updated perspective on the enduring question of why people dislike inflation, incorporating significant advancements in survey methodology that have occurred since the 1990s. I designed and conducted two new surveys on large, representative samples of the US population. The goal was to cover the perceived impacts of and reactions to inflation with simple but comprehensive questions. Considering inflation's impact on individuals in their varied economic roles—be it as consumers, workers, or asset owners—is crucial. Survey A contains detailed, closed-ended questions formulated in line with contemporary best practices to capture a spectrum of perspectives and actions. Survey B, on the other hand, consists of open-ended questions, allowing participants to express their thoughts freely. These questions are vital as they illuminate the nuanced views and convictions that might not fit within the predefined choices an economist could propose and that might be overlooked otherwise. Analyzing the responses to these questions on a broad scale via text analysis techniques enables the exploration of significant first-order concerns (Ferrario and Stantcheva 2022). Furthermore, by gathering detailed demographic data on participants in these large and representative samples, I am able to examine heterogeneities in attitudes and responses across different demographics, including income, political orientation, age, education, gender, and race.

The key findings can be summarized as follows: contrary to perceiving inflation as a mere yardstick or a unit of measure, individuals anticipate a variety of tangible adverse effects on both their personal financial situation and the economy at large. If there is a single and simple answer to the question, "Why do we dislike inflation?" it is because many individuals feel that it systematically erodes their purchasing power. Many people do not perceive that their wage increases sufficiently to keep up with inflation rates, and they often believe that wages tend to rise at a much slower rate compared to prices.

This perception of diminished living standards due to inflation is intensified by the observation that individuals rarely ascribe the raises they receive during inflationary periods to adjustments for inflation. Rather, they attribute these increases to job performance or career progression, particularly among those who have switched jobs during such periods.

In response to the perceived erosion of purchasing power, respondents report having to make costly and significant adjustments to their consumer behavior, such as reducing the quantity and quality of goods purchased or deferring purchases. Understandably, lower-income respondents report being most adversely affected, indicating that they have even postponed buying essential items to cope with the impact of inflation. Notably, very few respondents report accelerating their desired purchases or stockpiling in anticipation of further price rises.

Not surprisingly, given these perceived consequences, inflation triggers stress and emotional reactions. Another factor contributing to the aversion toward inflation is a sense of unfairness. All perceived impacts—whether experienced as consumers, workers, or asset owners—are felt more acutely by those with lower incomes who find themselves needing to make more significant adjustments across these dimensions as well. In line with this observation, there is a common belief that the incomes of higher-earning individuals increase more quickly than theirs during periods of inflation, suggesting a perception that inflation exacerbates inequality.

Why do individuals believe that wages do not increase as rapidly as prices? A primary reason is the conviction that employers and companies possess significant discretion in setting wages and tend to resist adjusting them upward in order to enhance their profit margins. There's a prevalent view that firms make strategic choices, with a more limited belief in market forces driving decisions.

When asked about the causes of inflation, people tend to blame the government and businesses. There is a clear partisan divide in the responses, with Republicans more likely to blame the government or Joe Biden, and Democrats more likely to blame businesses. This closely correlates with whom people feel angry at when they see prices rise, directing blame at businesses, the government, and the system in general.

Furthermore, people scarcely acknowledge any positive impacts from inflation. Consequently, only a minority of respondents believe in the trade-off between inflation and unemployment or associate inflation with enhanced economic growth (Binetti, Nuzzi, and Stantcheva 2024). The majority link inflation to adverse wider economic and political outcomes. Considering the numerous negative and scant positive perceived effects, many participants rank inflation as a top priority, ahead of other economic and social issues.

Despite shifts in the economic landscape, the core conclusions from the seminal study conducted by Shiller (1997) in the 1990s are still relevant today. But I also add some new findings, specifically exploring the many margins along which people report making costly adjustments and a range of emotions and attitudes toward inflation using a mix of open-ended text and structured questions. Furthermore, I highlight the distinct polarization in opinions on inflation based on political affiliation, along with varying attitudes and responses according to income level.

RELATED LITERATURE This paper contributes to several strands of the literature. First, it connects with studies on attitudes toward inflation or policies to combat price increases, primarily using survey methods. Shiller (1997) provided a first seminal contribution. Subsequent work has tried to characterize inflation aversion (Scheve 2004; Easterly and Fischer 2001; Howarth and Rommerskirchen 2017; Aklin, Arias, and Gray 2022; van Lelyveld 1999; Di Tella, MacCulloch, and Oswald 2001; Hofstetter and Rosas 2021; Ruprah and Luengas 2011; Hübner and Klemm 2015; Coles and Chen 1990; Jayadev 2008; Scheve 2003) and fairness concerns for firms' pricing behavior (Rotemberg 2005, 2011).

A series of recent papers relates most closely to the question of why people dislike inflation. Like the current paper, Jain, Kostyshyna, and Zhang (2022) find that respondents in Canada tend to associate higher inflation with worse labor market conditions. They also show that respondents do not think that wages adjust fully to inflation and that higher inflation expectations are associated with lower expected real spending growth. Hajdini and others (2022) show that an experimentally induced increase in inflation expectations is positively correlated with higher growth expectations, but the pass-through is relatively small at 0.2. Higher-income respondents are more likely to perceive a positive link between inflation and growth, similar to my findings about the less negative attitudes toward inflation among the better-off. Kamdar (2019) finds that people generally believe that an increase in inflation will be associated with an increase in unemployment, echoing my results.

The paper is also related to the large body of literature on inflation expectations, reviewed in Weber and others (2022). Coibion, Gorodnichenko, and Kamdar (2018) emphasize the importance of survey-based measures of inflation expectations, which are more accurate than traditional rational expectations approaches. Several papers study how expectations are formed, particularly focusing on personal experiences (Angelico and Giacomo 2019; Cavallo, Cruces, and Perez-Truglia 2017; D'Acunto and others 2019; D'Acunto and others 2021; Bruine de Bruin, van der Klaauw, and Topa 2011; Goldfayn-Frank and Wohlfart 2020; Malmendier and Nagel 2016). Binder, Janson, and Verbrugge (2023) study the anchoring of inflation expectations among professional forecasters.[1] Coibion, Gorodnichenko, and Weber (2022) examine how monetary policy communications shape inflation expectations.

An important contribution to survey methodology for inflation expectations is by Kim and Binder (2023), who show that repeat survey participants exhibit "learn-through-surveys" effects, whereby they adjust their forecasts and reduce their errors over time. Reassuringly, given the size of the pool of respondents and the nature of typical surveys done on the platform used in this paper, it is highly unlikely that respondents have been surveyed on inflation before.

Echoing my analysis of the perceived causes of inflation, recent work studies the narratives people have regarding inflation (Andre and others 2021; Andre and others 2022), with similar findings to mine along that dimension. I also study the behaviors adopted by households when there is inflation, which relates to the literature on behavioral changes induced by inflation expectations (Bachmann, Berg, and Sims 2015; Coibion and others 2023).

Finally, this paper is part of a broader research agenda to understand how people reason about economic phenomena and policies, following work on climate change policies (Dechezleprêtre and others 2022), trade policy (Stantcheva 2023b), inflation (Binetti, Nuzzi, and Stantcheva 2024), and tax policy (Stantcheva 2021).[2]

The rest of the paper is organized as follows. Section I presents the survey and sample. Section II provides results on people's definitions of and interest in inflation, and their perceived broader causes and consequences of inflation. Section III considers the personal impacts of and reactions to inflation as consumers, workers, and asset holders, as well as the

1. See Binder, McElroy, and Sheng (2022) on forecasters' subjective uncertainty, as well as Coibion and Gorodnichenko (2015) for a study of the same professional forecasters data that rejects the full-information rational expectations model and shows that the data are most consistent with a violation of the full-information assumption.

2. A lot of the data can be found on the website, Understanding Economics, https://understandingeconomics.org/#/.

emotional and psychological impacts. Section IV studies how respondents rank inflation relative to other economic and social issues and how they perceive the inflation-unemployment trade-off. Section V concludes.

I. Survey and Sample

I.A. Data Collection and Sample

I collected responses for two surveys between December 2023 and January 2024 on the survey platform Lucid. Lucid is a survey marketplace that pools together respondents from different panels, and respondents are rewarded based on the agreements with their survey panels (some in the form of points or perks on various partnering programs with hotels, stores, or airlines, others in the form of cash).

For the first survey, survey A, I collected a total of 1,500 responses; for the second survey, survey B, I collected 504 responses. For both surveys, I imposed quotas on age, income, gender, region, and race, as well as screening questions toward the start of the survey to filter out careless respondents.[3]

Table 1 compares the characteristics of our sample to the US population. The samples are, by construction, closely representative along the targeted margins. For nontargeted margins, the samples match quite well for family structures, the share employed, the share Republican, and the share having voted for Biden versus Trump in 2020. As with almost all online surveys, there is some oversampling of college-educated and unemployed respondents (Stantcheva 2023a). The sample share of Democrat respondents relative to the share of independents is also larger than in the US population, although the voting shares for 2020 match much more closely.

I.B. Survey Structure

Survey A contained closed-ended questions. The full questionnaire can be found in online appendix A.4. The survey covered the following topics: definition of inflation, information about past inflation and expected inflation, personal impacts and reactions to inflation, and policy views related to inflation. This survey took on average 32 minutes to complete (median 27 minutes).

Importantly, these survey questions were designed with the clear intention of *not* priming respondents to answer in a given way. For instance, even if economic theory or evidence says the direction of an effect is unambiguous,

3. Those respondents were immediately screened out of the survey and not allowed to complete it.

Table 1. Sample Representativity

	Survey A	*Survey B*	*US population*
Targeted characteristics			
Male	0.48	0.50	0.49
Female	0.51	0.50	0.51
18–29 years old	0.23	0.22	0.23
30–39 years old	0.21	0.21	0.21
40–49 years old	0.19	0.20	0.19
50–59 years old	0.19	0.18	0.19
60–69 years old	0.18	0.19	0.18
$0–$19,999	0.14	0.15	0.13
$20,000–$39,999	0.16	0.15	0.16
$40,000–$69,999	0.20	0.20	0.20
$70,000–$99,999	0.15	0.15	0.15
$100,000–$124,999	0.08	0.10	0.09
$125,000+	0.26	0.25	0.26
White	0.68	0.64	0.60
African American/Black	0.12	0.13	0.13
Hispanic/Latino	0.13	0.16	0.19
Asian/Asian American	0.03	0.04	0.06
Northeast	0.19	0.19	0.18
South	0.37	0.39	0.37
Midwest	0.21	0.20	0.21
West	0.23	0.22	0.24
Nontargeted characteristics			
Married	0.49	0.48	0.52
Single	0.37	0.35	0.35
Separated/divorced	0.10	0.13	0.12
Widowed	0.03	0.04	0.02
Has children	0.59	0.64	0.40
Less than high school	0.03	0.04	0.09
Less than four-year college	0.51	0.53	0.55
Four-year college/master's	0.40	0.33	0.32
Professional degree	0.06	0.11	0.03
Employed	0.65	0.73	0.70
Unemployed	0.09	0.07	0.03
Republican	0.28	0.32	0.26
Democrat	0.38	0.34	0.25
Independent and others	0.34	0.34	0.47
Voted in 2020 presidential election	0.80	0.81	0.61
Voted for Biden in 2020 presidential election	0.56	0.53	0.51
Voted for Trump in 2020 presidential election	0.40	0.43	0.47
Sample size	1,500	504	

Source: Author's surveys and IPUMS-CPS-ASEC.

Note: The table displays statistics for the overall US population, as compared to the samples of respondents in surveys A and B. Summary statistics for the US population are constructed using IPUMS-CPS-ASEC data for 2022. Targeted characteristics refer to the ones on which we impose quotas in our survey to match the overall US population. Quotas are not set for the nontargeted characteristics.

the question still features a bilateral scale allowing respondents to take a stand on the direction. The questions are balanced, neutral, and clarify terms as needed, following the best practices outlined in Stantcheva (2023a).

Survey B focused on open-ended questions. It covered topics such as respondents' perceived causes and consequences of inflation, emotional reactions to inflation, and personal impacts. The full questionnaire can be found in online appendix A.5, and the survey took on average 14 minutes (median 11 minutes) to complete.

The responses to open-ended questions are valuable: they provide us with respondents' views before they are primed to think in any particular direction by the surveyor. They can convey issues that we might otherwise miss. To analyze these answers, I create topics defined by lists of keywords and categorize answers depending on whether they contain the keywords associated with the topic (Ferrario and Stantcheva 2022). A given answer may contain more than one topic, which is why some respondents may be reflected in one or more categories. Furthermore, a (typically) small share of responses are not classified because they do not fit into a clear category or do not answer the question. As a result, the categories do not systematically add to 100 percent. I also chose to report the answers as they were written by respondents when providing examples, which means they may contain typos and errors. Online appendix A.3 provides example answers for each question and category.

In both surveys, I occasionally used a question from Shiller (1997) when it is particularly interesting to make an exact comparison between the views in 1996 and those today. Nevertheless, I rephrased most of the questions to be more balanced and neutral, and I added extensive new questions to better understand people's reasoning.

I.C. Paper Organization

Throughout the paper, I will draw on responses from both surveys, specifying each time whether the question under consideration is open-ended or closed-ended. Figures A14–A18 in the online appendix depict the raw word clouds from the open-ended questions.

In some analyses, I will highlight the heterogeneity in views by income, with groups defined as those in the lower third of the income distribution of respondents (income below \$40,000) and those in the upper third (income above \$125,000). In others, the heterogeneity by political leaning is more interesting to showcase. I also systematically show the sample average. Online appendix A.2 contains the complementary figures that are not shown here.

Furthermore, tables A2–A23 contain detailed regression results, where all outcomes shown in the figures are regressed on the full set of individual characteristics. These tables show that the patterns highlighted in the main text figures also hold when controlling for detailed individual covariates, and they highlight further heterogeneities by education, age, race, or employment status. Due to space constraints, I cannot discuss these other heterogeneity patterns at length here.

II. Understanding, Expectations, and Interest in Inflation

II.A. Inflation Definition

The first set of results relates to people's basic understanding of inflation.

First, it is instructive to ask people about their definition of inflation, in their own words. Table 2 shows example responses to this open-ended question. Around half of all respondents give a relatively correct response. In their own words, "Inflation is the price of things going up," "I describe inflation as an increase in prices across the country," "A rise in the general price of goods." Very few respondents provide the exactly correct definition of inflation, and there are clearly some difficulties with the formal definition, whereby people tend to add extra clauses or conditions to it.

On the contrary, 44 percent of respondents give relatively incorrect answers, with examples such as "The hiking of prices of consumer goods to offset the country's debt due to elites over spending and throwing money away," "Price gouging, especially for the greedy, by raising prices so high, that almost everything is too expensive," "Inflation is when everything gets so expensive. You can't afford it no matter how hard you work," "Inflation to me is where the cost of living rises above affordable means for the majority of the people," and "Over priced everything."

However, in simple, concrete examples, many more people are able to correctly estimate the inflation rate. I asked respondents two short knowledge questions: the first told them the price of a good today, gave them an annual inflation rate, and asked them to compute the price of the good one year from now. Table 3 shows that 85 percent of respondents did this correctly. Conversely, the second question gave them the current price and the price one year from now and asked them to compute the inflation rate; 82 percent of people got this right. Therefore, simple exercises may, understandably, not reflect people's true grasp of the underlying concept.

I included in the survey an interesting question from Shiller (1997), asking people whether they agreed with a characterization of inflation as a "sort of measurement thing/yardstick and little more." Both in 1996 and

Table 2. A Closer Look at Definitions of Inflation

Relatively correct answers (52 percent)	*Relatively incorrect answers (44 percent)*
Inflation is the price of things going up.	The hiking of prices of consumer goods to offset the country's debt due to elites overspending and throwing money away.
I describe inflation as an increase in prices across the country.	Inflation is when everything gets so expensive. You can't afford it no matter how hard you work.
Inflation is when the price of goods go up based on the economy.	Inflation to me is where the cost of living rises above affordable means for the majority of the people.
Inflation is when the price of things go up over time. This can be attributed to specific events that cause the rise of pricing.	Price gouging, especially for the greedy, by raising prices so high, that almost everything is too expensive.
A rise in the general price of goods.	Overpriced everything.
Inflation is a rise in prices, which can be translated as the decline of purchasing power over time.	The price of goods keeps increasing but our income doesn't.
The rise of prices for goods and services.	Not being able to afford to live.
Inflation is the general increase in the prices of goods and services in an economy over a period of time.	To me, inflation is when the economy is more than just hurting. It's when it's too tough just to keep positive.
Inflation is the increase of prices of goods.	Increase in demand.
Inflation is the rising cost of prices across multiple industries including food, electronics, and automobiles.	Goods and services are priced high. The costs are inflated.

Source: Author's surveys.

Note: This table offers ten examples of correct and incorrect answers to the question, "How would you define inflation in your own words?" Note that 4 percent of respondents answered without giving any definition. Answers are reported as they were written by respondents when providing examples, which means they may contain typos and errors.

today, a minority of people (40 percent) agree with this description. This disagreement will not be surprising in light of the range of (real) consequences people expect from inflation, which I present below. I provided respondents with a definition of inflation before moving to the actual questions about it.

II.B. Past Inflation and Inflation Expectations

Turning to knowledge of the past inflation rate and inflation expectations, 92 percent of people think that there has been inflation (as opposed to deflation or no change in prices) over the last twelve months. Nearly three-quarters of respondents expect inflation to continue over the next year,

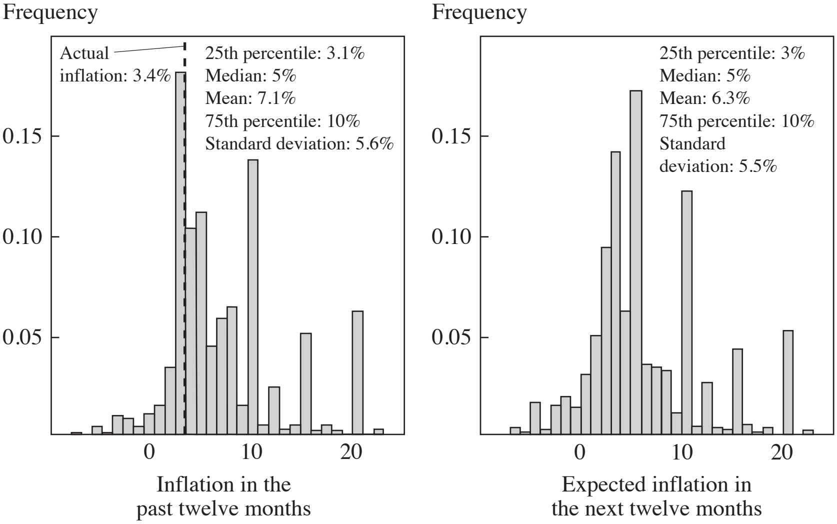
Table 3. Understanding and Importance of Inflation

	Share of respondents giving each answer
Understanding of inflation	
Correct future price given inflation rate	0.85
Correct inflation rate given future price	0.82
Agree with the definition of inflation as a "sort of measurement thing and little more"	0.40
Over the last twelve months	
Inflation	0.92
Deflation	0.04
No change in prices	0.04
Over the next twelve months	
Inflation	0.72
Deflation	0.09
No change in prices	0.19
Items which experienced the most substantial inflation in past twelve months	
Food	0.59
Gas	0.19
Rent	0.15
Utilities	0.06
Main source of news about inflation	
Social media	0.47
Newspapers	0.62
Television	0.76
Radio	0.37
Most influential source when thinking about future inflation	
News reports	0.13
Official statistics	0.20
Recent price changes of my purchases	0.65
Advice from friends and family	0.02
Attention for inflation updates	
Find important staying up to date on current and future inflation	0.71
Increased attention toward inflation in last two years	0.82
Sample size	1,500

Source: Author's surveys.

Note: The third variable is an indicator equal to one if the respondent somewhat to strongly agrees with the statement. Respondents could select several main sources of news about inflation. The indicator "Find important staying up to date on current and future inflation" is equal to one if the respondent finds being updated very to extremely important. The indicator "Increased attention toward inflation in last two years" is equal to one if the respondent increased attention somewhat to a lot. For more details on the questionnaire, see online appendix A.4.

Figure 1. Distribution of Estimates of Past and Expected Future Inflation (Censored)



Source: Author's surveys and Bureau of Labor Statistics.

Note: Data for actual inflation from December 2022 to December 2023 retrieved from the Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers, all items in US city average, not seasonally adjusted, accessed at https://www.bls.gov/regions/mid-atlantic/data/consumerpriceindex historical_us_table.htm. Samples of perceived inflation rates and expected future inflation rates are censored at –10 percent (excluding, respectively, 0.6 percent and 0.7 percent of the sample) and 25 percent (excluding, respectively, 7.7 percent and 5 percent of the sample).

while almost 20 percent expect a stabilization of prices. Figure 1 plots the distribution of past and expected inflation rates across respondents. While actual inflation over that period was 3.4 percent, the median expectation is a bit higher at 5 percent, and the mean is much higher at 7.1 percent. Median expected inflation over the next twelve months is identical to the median past expectation at 5 percent and the mean is 6.3 percent.

Online appendix table A2 correlates the perceived past and expected inflation with various socioeconomic characteristics. There are some striking differences in perceptions and expectations across respondents. High-income respondents perceive around 3 percentage points lower past and expected inflation. Republican, female, and Black respondents think inflation has been higher in the past and have higher inflation expectations for the coming year.[4]

4. Bruine de Bruin and others (2010) find that inflation expectations are higher for non-white, less-educated, and lower-income respondents. Unlike us, they find a significantly positive effect of age but no effect of gender on inflation expectations.

Table 3 also reflects the items that people believe have experienced the most substantial inflation over the past year: food leads the ranking, followed by gas, rent, and utilities.

I want to emphasize that there are many issues with how inflation is measured—due to unavoidable assumptions that have to be made—so that official measures might not reflect the experience of specific groups. Two important measurement issues are, first, inflation inequality and, second, the way housing and financing costs are taken into account. These will introduce a discrepancy between people's experienced inflation and official inflation statistics. As a result, perceived and expected inflation—and by extension, perceived living costs and real wage growth—might deviate from official numbers.

Inflation inequality means that inflation might affect households differently because of the basket of goods they consume (Jaravel 2021; Atkin and others 2024; Argente and Lee 2021; Cavallo, 2024; Wimer, Collyer, and Jaravel 2019; Jaravel and Olivi 2021; Kaplan and Schulhofer-Wohl 2017). In general, lower-income households, which spend a larger share of their budget on food, gas, rents, and necessities are likely to be more affected. Furthermore, experienced inflation will differ across space in the United States. This inflation inequality means that real wage growth might also not be accurately computed for households at different points in the income distribution or living in different places in the United States. Related to housing and financing costs, a recent paper by Bolhuis and others (2024) notes that consumers consider financing costs—for mortgages, auto loans, and other personal loans—and leasing costs to be part of the cost of living. Yet, these costs are not part of the current Consumer Price Index (CPI). Therefore, the current measure of inflation does not capture the effective costs that are facing potential home buyers and those relying on financing instead of cash purchases. Bolhuis and others (2024) show that a modified CPI taking these costs into account exhibits much higher inflation in the recent period.

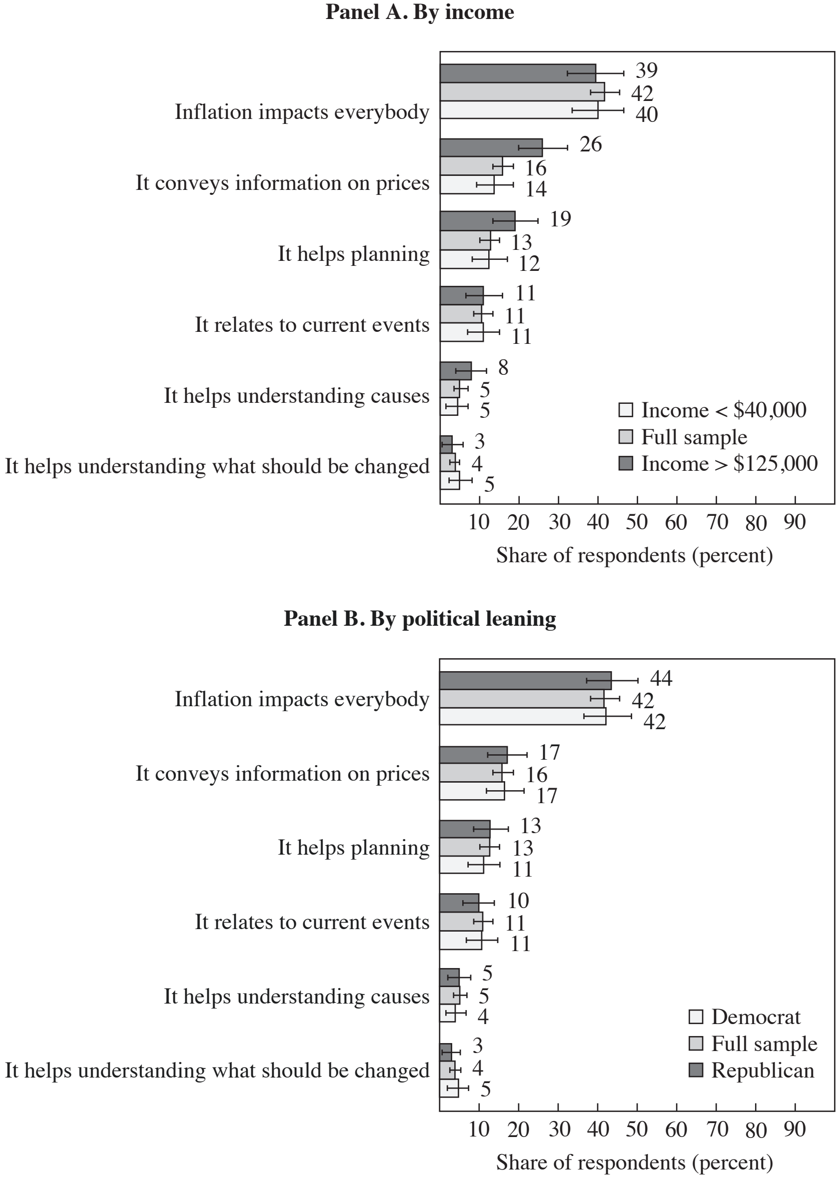
These measurement issues imply that people's perceptions may accurately reflect their true experience even if they are not in line with official statistics.

II.C. Interest in Inflation and Sources of Information

Table 3 shows that 71 percent of respondents find it "extremely important" to stay up to date on inflation, and 82 percent report that their attention to inflation news has increased over the last two years.

Why are people interested in inflation? Figure 2 shows the answers to the open-ended question from survey B, which reads, "Some people

Figure 2. News on Inflation Is Interesting Because . . . [Open-Ended Text]



Source: Author's surveys.

Note: The figure reports the share of respondents whose answers belong to each category with 90 percent confidence intervals. The question is, "Some people think that news about inflation is boring and technical stuff that they can't relate to. Can you explain to them why they should find it interesting?" For each category, I report two example answers in online appendix A.3.1. Seven percent of respondents answered they were not interested in news about inflation.

think that news about inflation is boring and technical stuff that they can't relate to. Can you explain to them why they should find it interesting?" The most common answer, across income and political groups, is that inflation affects everyone (example answers include "Because it affects everyone's lives" or "It affects everyone's cost of living"), followed closely by the fact that news conveys information about prices (with example answers such as "Could be an indication of future price increases").

The main sources of formal news about inflation reported are television, followed by newspapers, social media, and, finally, radio. Yet, news does not appear to be the main driver of expectations. When I ask people what source is most influential for them when they form their views about future inflation, it appears that people by far infer the most information from their recent purchases and the price changes they witness when shopping (see table 3). Around one-fifth rely on official statistics, and only 13 percent rely on news reports.

II.D. Perceived Causes of Inflation

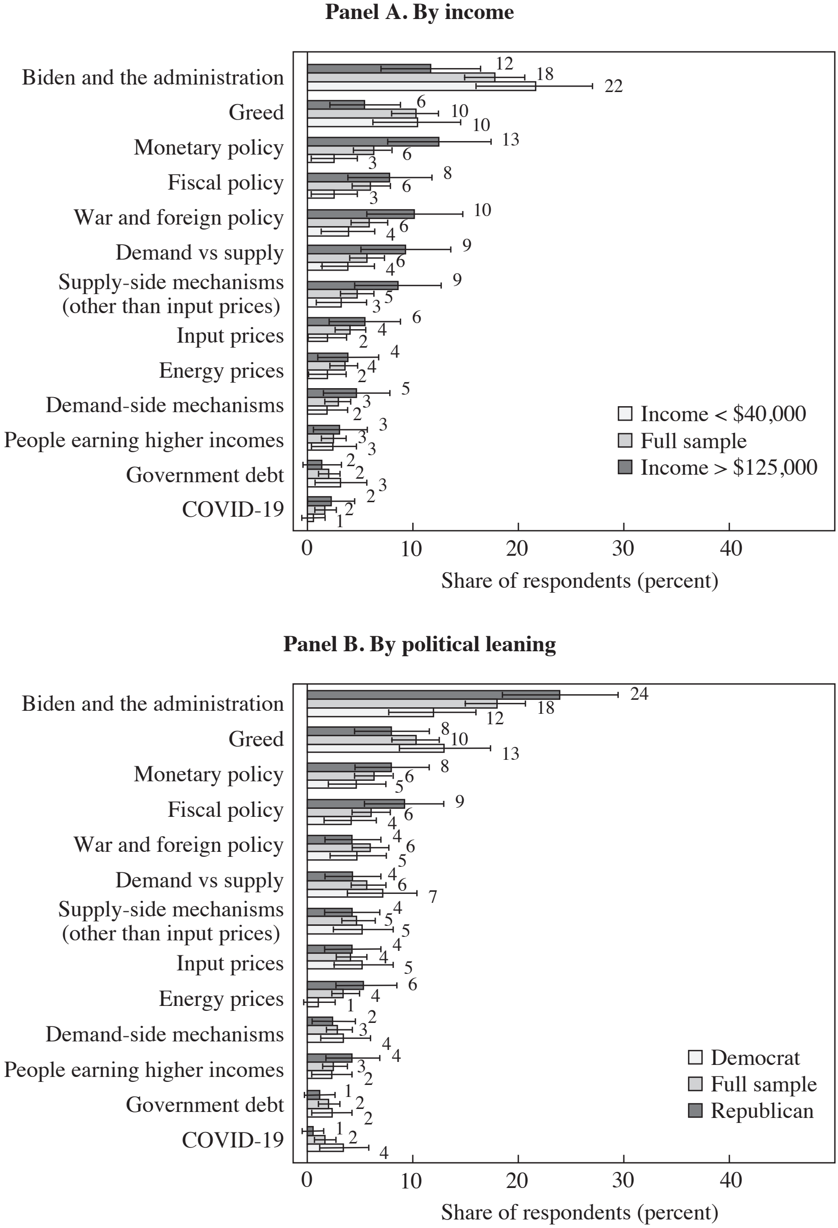
To continue gauging respondents' core understanding of inflation, I also ask them open-ended questions about the consequences and causes of inflation.[5]

Starting with the causes of inflation, figure 3 shows that, when respondents are asked in an open-ended way without priming them about specific causes, there is a large variety of causes mentioned. The most common one is Biden and the administration ("I think it has to do with Joe Biden," "Joe Biden's policies for this round of inflation"), followed by greed ("I believe the sole reason is greedy corporations who care more about their bottom line than actually helping people," "I think is some cases it is price gouging. When you know people depend on a product you want to see at what price are they still willing to pay for it"). There is a clear partisan divide in the perceived importance of these two main causes. Democrats are much more likely to talk about greed, while Republicans more frequently point to Biden and the administration.

Monetary policy ("Too much money injected into the market by the Fed," "Low interest rates") is especially mentioned among higher-income respondents (13 percent of them), but only among 3 percent of lower-income ones. Online appendix table A4 shows it is also more commonly mentioned among college-educated respondents.

5. A more in-depth analysis is in Binetti, Nuzzi, and Stantcheva (2024).

Figure 3. High Inflation Is Caused by . . . [Open-Ended Text]



Source: Author's surveys.

Note: The figure reports the share of respondents whose answers belong to each category with 90 percent confidence intervals. The question is, "When inflation gets very high, what do you think is the reason?" For each category, I report two example answers in online appendix A.3.4.

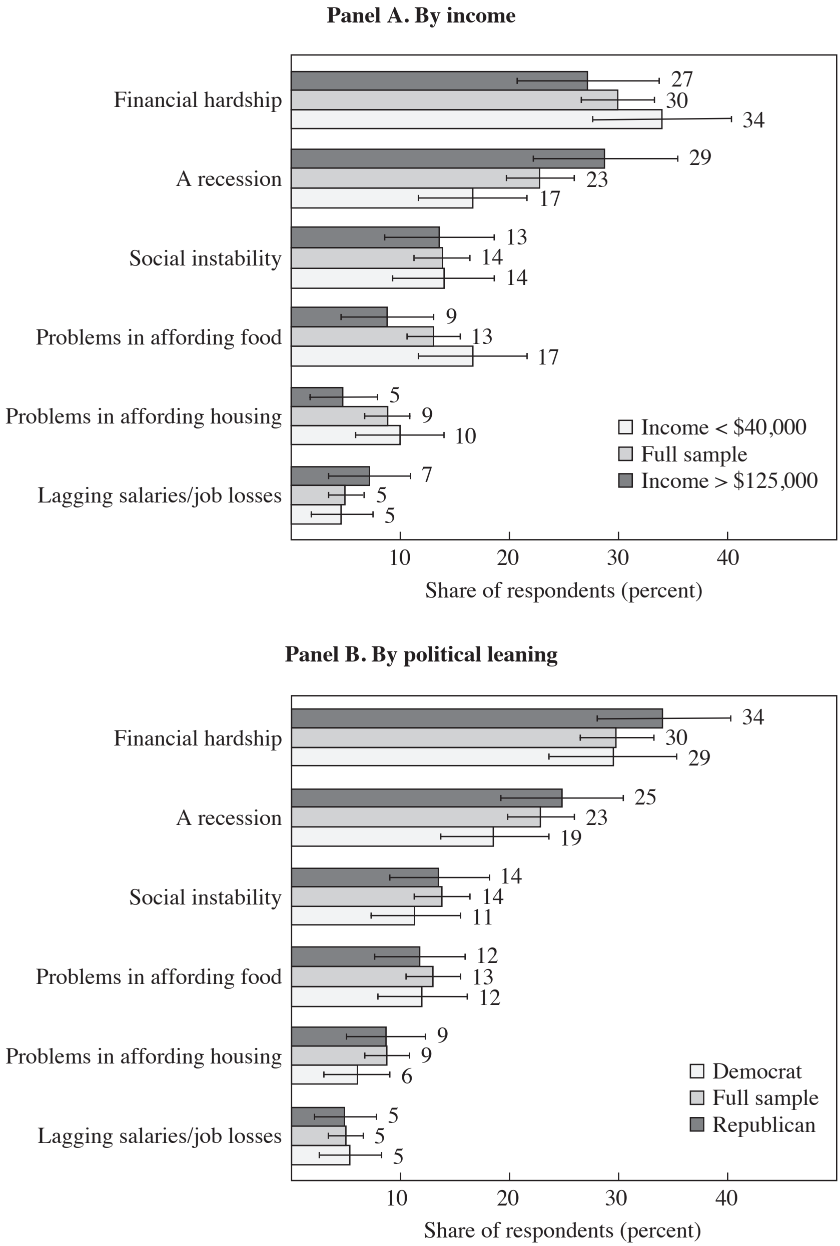
Some respondents (10 percent or fewer in all cases) also mention fiscal policy ("Government overspending is one principal reason," "Tax breaks for the rich and poor budgeting"), war and foreign policy ("I think it's because of war," "It can be many factor, but the main factor is related to trade with other countries. When sanctions are in place, imports are reduced therefore limiting our supply of certain products"), demand versus supply ("I think the reason is supply and demand—the demand is high and goods are scarce," "Because there is a problem with supply and demand"), supply-side mechanisms, other than input prices ("Because we have a shortage on supply," "Supply chain issues"), input prices ("Companies raising their manufacturing costs," "Costa of things and materials to make them") specifically, energy prices ("Because gas prices, rises, losses rises"), and to a lesser extent, demand-side mechanisms ("Devaluation of dollar and excessive demand of products," "I think it's because the high demand of a product"). Perhaps surprisingly, very few people mention COVID-19 as a main cause.

II.E. Perceived Consequences of Inflation

ANTICIPATED POSITIVE AND NEGATIVE CONSEQUENCES OF INFLATION Figure 4 shows the responses to the question, "If inflation increases too much, what do you worry might happen?" The most common answer is related to financial hardship, with examples such as "I won't be able to afford essential items" or "That we can no longer afford our basic human rights to live." The share of respondents mentioning this issue is larger among lower-income respondents and Republicans. Other consequences mentioned in order of importance relate to the risk of a recession ("We might go into another great Depression," "Financial crash"), social instability ("Theft and crime are rising because of it"), problems in affording food ("That food prices will be so high that I could barely feed my family," "That it might go too high that people can't afford food"), problems in affording housing ("That I will be homeless," "I can't afford anything and lose my home"), and lagging salaries/job losses ("I am worried it might affect wages. If wages are not keeping up with inflation, we would be able to buy less with our paycheck," "People will start losing their jobs"). All these concerns are more widespread among low-income respondents with the exception of the general recession risk, which is more common among high-income respondents.

Do respondents perceive any positive impacts from inflation at all? Figure 5 shows that the answer is generally mixed: 60 percent of low-income respondents (as compared to 31 percent of high-income ones) believe there are no positive impacts of inflation at all. The share is also

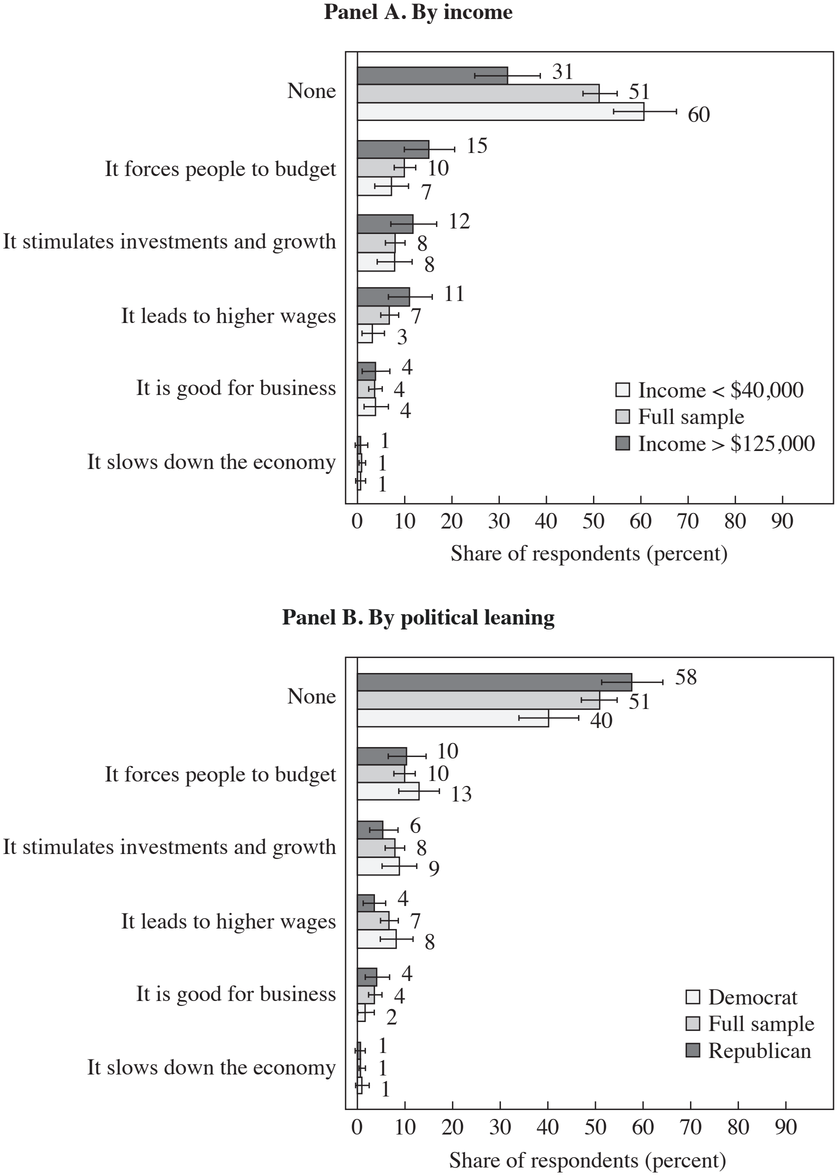
Figure 4. If Inflation Increases Too Much, I Worry about . . . [Open-Ended Text]



Source: Author's surveys.

Note: The figure reports the share of respondents whose answers belong to each category with 90 percent confidence intervals. All the shares reported here are unconditional. The question is, "What are you worried might happen?" For each category, I report two example answers in online appendix A.3.2.

Figure 5. A Positive Impact of Inflation Is . . . [Open-Ended Text]



Source: Author's surveys.

Note: The figure reports the share of respondents whose answers belong to each category with 90 percent confidence intervals. The question is, "What do you think could be the positive effects of inflation, if any, on people's economic and financial situation?" For each category, I report two example answers in online appendix A.3.3. Twenty-one percent of the answers are not reported in the figure since they either mention a benefit that appears only once or twice or do not answer the question.

higher among Republicans than Democrats (58 percent compared to 40 percent). The main potential positive effect perceived is that it will force people to budget ("It will show people how to manage their money," "It forces people to budget") or will lead to higher wages. Consistent with what we will see below on the perceived unemployment-inflation trade-off, only very few respondents (8 percent on average) believe higher inflation can lead to higher growth. Higher-income respondents are more likely to report any of the potential positive impacts listed in the figure. The absence of a trade-off between inflation and economic activity and the fact that inflation is considered a "bad" that need not happen are explored in-depth in Binetti, Nuzzi, and Stantcheva (2024).

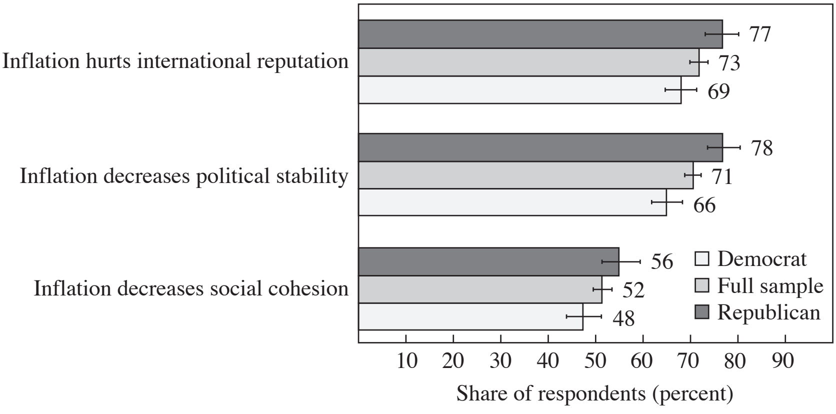
SOCIAL AND POLITICAL CONSEQUENCES OF INFLATION People's heightened interest in inflation becomes even more understandable when considering the far-reaching consequences people anticipate, above and beyond the personal impacts. Figure 6 shows that close to three-quarters of all respondents believe that "inflation hurts international reputation" and "decreases political stability." Views are more evenly split when it comes to decreasing social cohesion.[6] Negative perceived consequences are somewhat more salient among Republicans than Democrats, but as online appendix figure A3 shows, there is no systematic pattern by income.

Shiller (1997) asks a much starker question about whether there can be political and economic chaos if inflation gets out of control, which three-quarters of respondents agree with. But it seems that today, that same share agree also with less stark statements such as the ones above. The share who believe that inflation can hurt international prestige is similar in our sample and in Shiller (1997). Perhaps the recent episode of inflation has brought back inflation concerns that might previously have been assuaged by a long period of low inflation.

THE PERCEIVED LINKS BETWEEN INFLATION AND WAGES I also ask respondents about their theory of how inflation affects wages, keeping the question very similar to that in Shiller (1997). Three alternative theories are offered (figure 7). "Inflation will increase my employer's profits, but she will not feel the need to increase my pay" by far reflects the most held view with, on average, 51 percent of respondents selecting it. The share is higher at 54 percent among lower-income respondents than among higher-income ones. The share of all respondents who hold this belief is strikingly

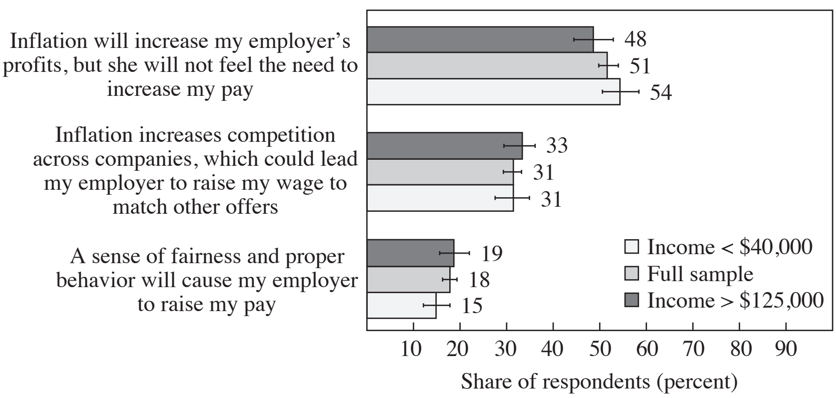
6. Here again, I do not prime respondents about the direction of the effect and provide bilateral answer options.

Figure 6. Perceived Social and Political Consequences of Inflation



Source: Author's surveys.
Note: The figure reports the share of respondents whose answers are reflected by the statements listed alongside 90 percent confidence intervals. For more details on the questionnaire, see online appendix A.4.

Figure 7. Theories about Inflation and Wages



Source: Author's surveys.
Note: The figure reports the share of respondents selecting each theory alongside 90 percent confidence intervals. For more details on the questionnaire, see online appendix A.4.

similar to that in Shiller (1997), conditional on respondents answering the question, reflecting the widely held perception that employers' preferences determine wages to a large extent, rather than market forces. Around one-third of respondents across all income groups hold the view that "inflation increases competition across companies, which could lead my employer to raise my wage to match other offers." Finally, a smaller share, between 15 percent for lower-income respondents and 19 percent for higher-income ones, believe most in the theory that "a sense of fairness and proper behavior will cause my employer to raise my pay."

People's views about the link between inflation and wages may depend on the type of firm considered. To test this, I designed a series of questions about small and large firms. The results, reported in figure 8, show that, on balance, people believe that only a few or almost no firms will actually adjust wages to inflation, especially among small firms.

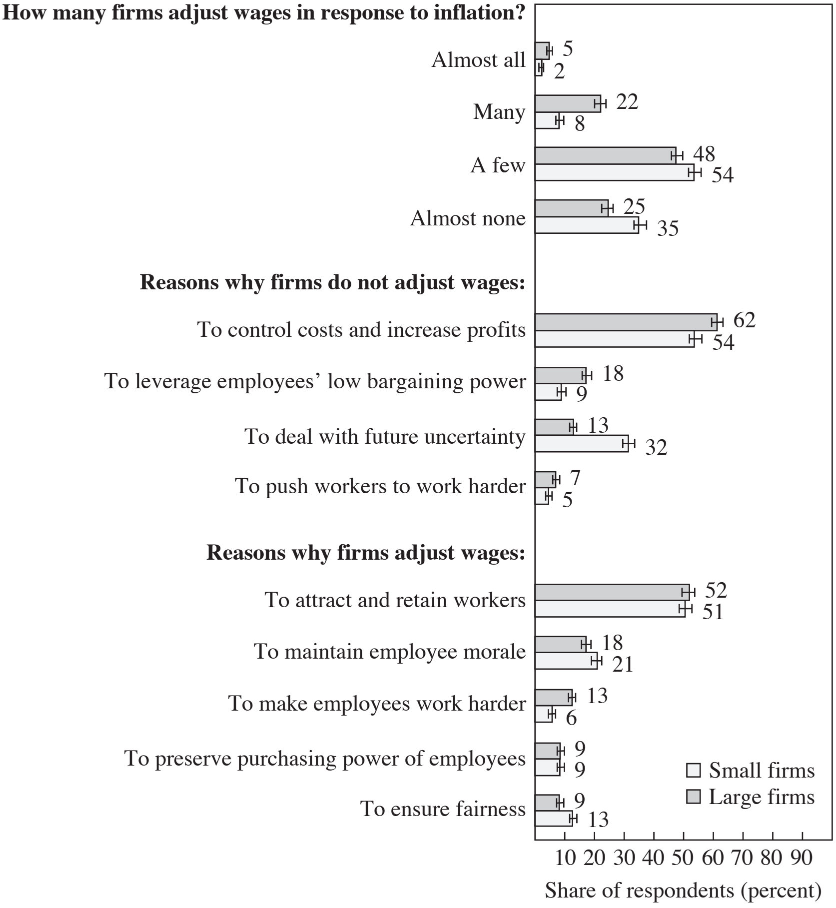
Most firms—large and small—are perceived to avoid adjusting wages to control costs and increase their profits (already echoing the notion of greed often heard in the news). Conditional on not adjusting wages, respondents are more likely to say that large firms are trying to leverage employees' low bargaining power, while small firms are dealing with future uncertainty. The main reason for adjusting wages, in people's views, is to attract and retain workers, followed by maintaining employee morale.

III. Personal Impacts of and Reactions to Inflation

Inflation can impact people in several roles: as consumers, as workers, and as asset holders. Before diving into people's experienced impacts along these specific dimensions, it is worth considering their answers to the open-ended question, "What were the most important impacts of inflation on your life?" shown in figure 9 (see also the word cloud in figure A16 in the online appendix). It is clear that the first-order concerns of most people are around the cost of living and affordability. Nearly one-third of respondents mention the cost of living in general, and over one-third mention either food affordability or gas affordability. Fewer people worry about the reduction in the value of their savings. Concerns about job losses are less of a first-order.

In this section, I consider people's various roles (consumers, workers, asset holders) in turn and study the perceived impacts of inflation and their responses to it. On this issue, the major heterogeneities are by income, which is why many of the figures focus on this dimension. For the figures by political leaning, see online appendix A.2.

Figure 8. Wage Adjustment in Small versus Large Companies

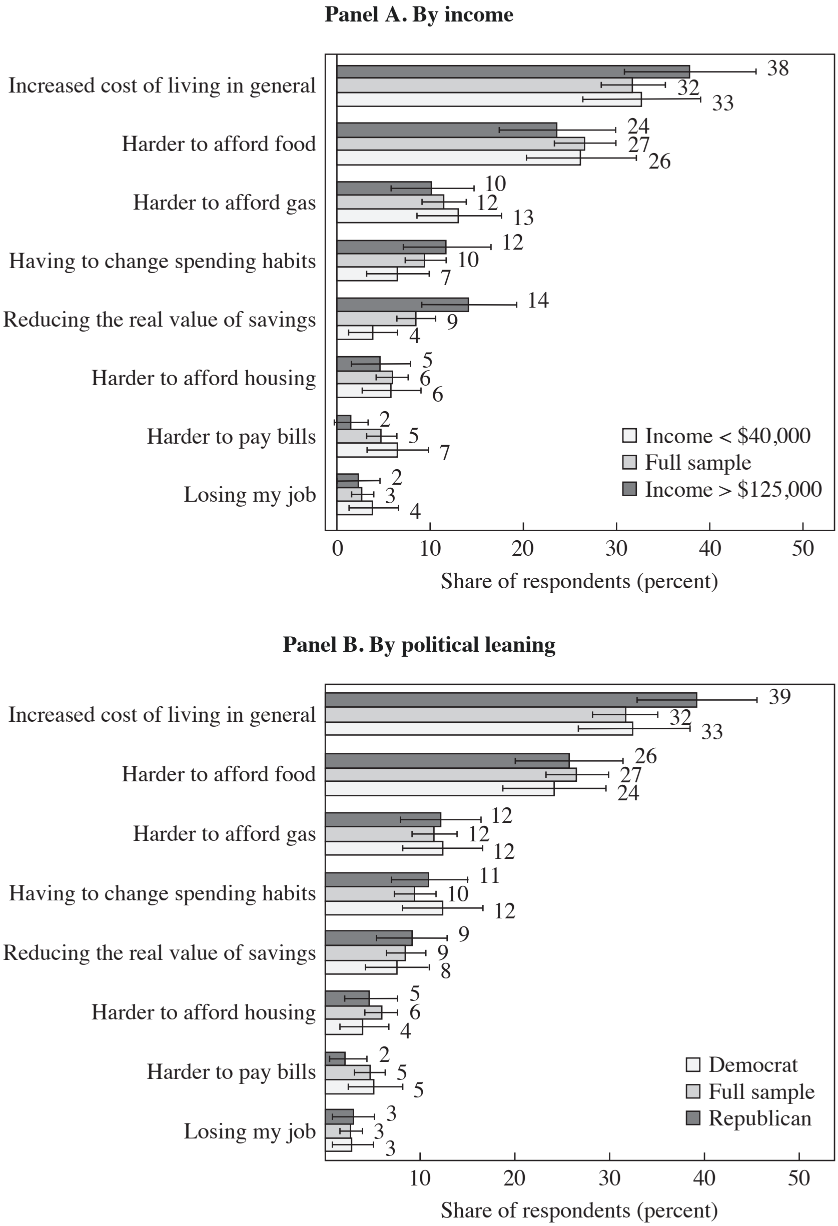


Source: Author's surveys.
Note: The figure reports the share of respondents whose answers are reflected by the statements listed alongside 90 percent confidence intervals. For more details on the questionnaire, see online appendix A.4.

III.A. As a Consumer

IMPACTS To better understand how people believe they experience impacts as consumers, figure 10 plots the distribution of answers to various questions. Consistent with the open-ended questions above, nearly three-quarters of the sample believe their purchasing power has decreased, which is remarkably similar to the 77 percent found by Shiller (1997) in response to this same question. This share is significantly higher among lower-income respondents in my sample.

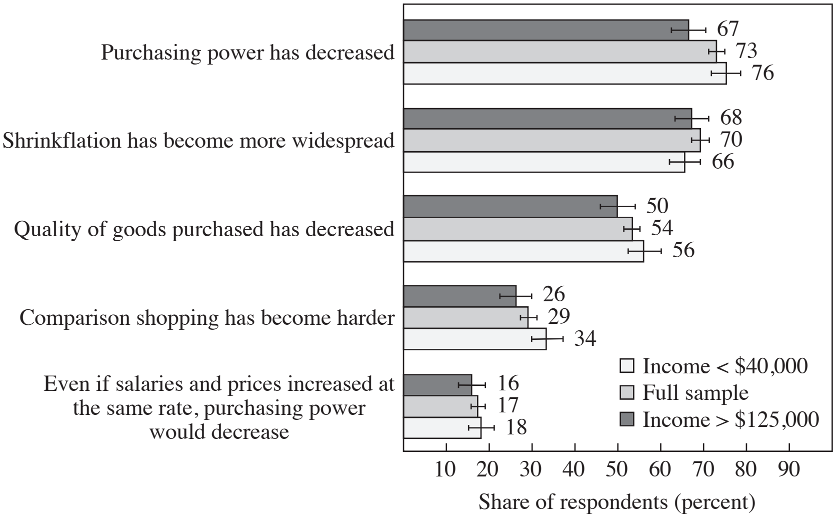
Figure 9. The Most Important Impact of Inflation on My Life Has Been . . . [Open-Ended Text]



Source: Author's surveys.

Note: The figure reports the share of respondents whose answers belong to each category with 90 percent confidence intervals. The precise question is, "What were the most important impacts of inflation on your life?" For each category, I report two example answers in online appendix A.3.5.

Figure 10. Inflation Impacts as a Consumer



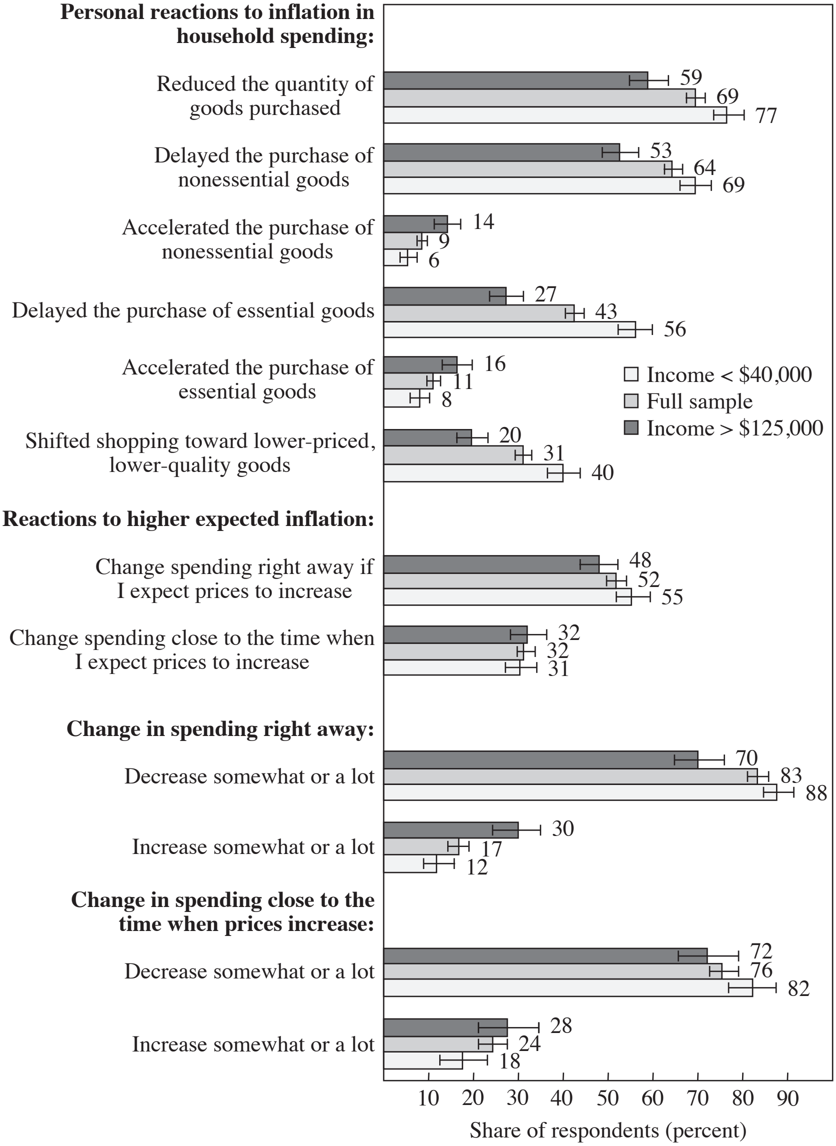
Source: Author's surveys.
Note: The figure reports the share of respondents whose answers are aligned with the statement listed alongside 90 percent confidence intervals. For more details on the questionnaire, see online appendix A.4.

Around 70 percent of respondents also believe that "shrinkflation," defined as a good having the same price but with reduced quality or quantity, has become more widespread. Less common (for around half of respondents) is the perception that the quality of goods purchased overall has decreased. Around one-third of respondents think that comparison shopping has become harder, which is higher than the 7 percent reported for a similar, but not identical, question in Shiller (1997), which suggests that price comparisons have become harder despite today's technologies.

REACTIONS How do people react when faced with these consequences of inflation? Figure 11 depicts a range of potential consumer reactions. Among lower-income respondents, a large share reduce the quantities of goods they purchase (77 percent) and delay the purchase of nonessential goods (69 percent). Around 56 percent report delaying the purchase of even essential goods.[7] A substantial share also report shifting toward

7. Note that these questions do not prime respondents about the direction: the questions let the respondents select between accelerating and delaying purchases.

Figure 11. Personal Reactions to Inflation as a Consumer



Source: Author's surveys.

Note: The figure reports the share of respondents whose answers are reflected by the statement listed alongside 90 percent confidence intervals. In the second set of bars, I show respondents' answers to the question of how they would change their spending if they expected prices to increase in the next year. Answers in the third and fourth set of bars are conditional on having chosen either "change in spending right away" or "when prices increase," respectively. For more details on the questionnaire, see online appendix A.4.

lower-priced and, accordingly, lower-quality goods. The numbers are much lower among high-income respondents, but nevertheless, a small majority says they will reduce purchases and delay nonessential ones.

Very few respondents report that they would accelerate the purchases of either essential or nonessential goods. The share is somewhat higher among high-income respondents (15 percent on average for these two categories) than for low-income respondents (7 percent on average), suggesting that high-income respondents might be more able to buy ahead of time.

I also ask respondents what they would do if they expected prices to increase in a year. More than half of all respondents report that they would start adjusting their spending right away, and conditional on doing so, they mostly report starting to decrease their spending at least somewhat. Nearly one-third of respondents instead say they will start adjusting closer to the time of the price change, but similarly, mostly again to decrease their spending. Thus, interestingly, respondents do not report trying to accelerate their purchases or create a stockpile either during an episode of inflation or in the (hypothetical) scenario of higher future inflation.

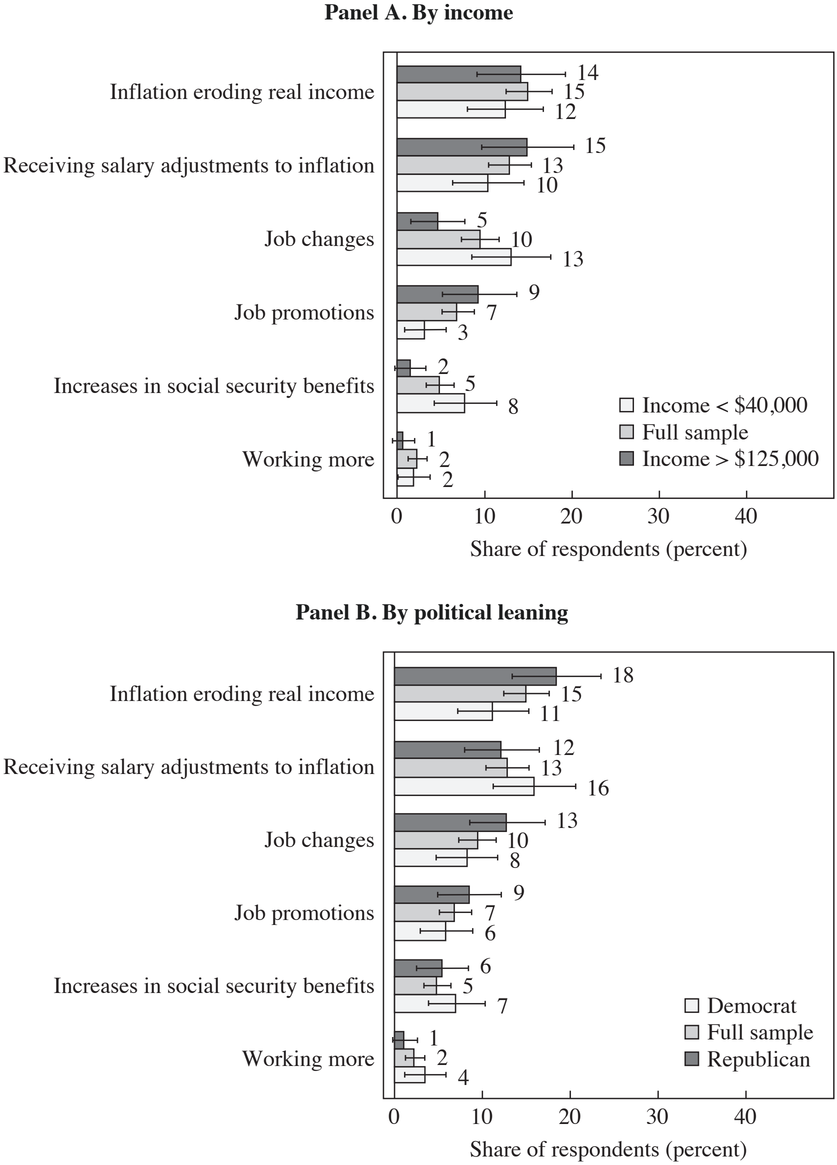
III.B. As a Worker

I also elicited people's views about how inflation affects them as workers and how they have responded to it.

IMPACTS First, to avoid priming respondents, I ask an open-ended question in survey B: "Think about how much your income (measured in dollars per month) went up (or down) in the past five years. What do you think are the most important factors that account for the change in your income?" The results, shown in figure 12, indicate that nearly one-third of respondents believe inflation is a primary cause of their income changes, and this group is split into equal shares between those who think inflation has eroded their real income ("Our income went up but we have far less money because of inflation," "The cost of living has gone up and wages have remained the same") and those who believe they have received income increases as adjustments for inflation ("My income has risen due to negotiated cost of living adjustments that are applied across the board to employees where I work" or "When I get a cost of living increase, it is because of inflation makes it necessary"). Only 10 percent or fewer of respondents believe wage changes were mainly due to job changes or promotions at work.

Figure 13 summarizes the key findings from closed-ended questions related to wage impacts. First, respondents are asked how long it would take for their wage to catch up if inflation doubled. About half of the sample believe it will take more than one year. Although only about one-quarter

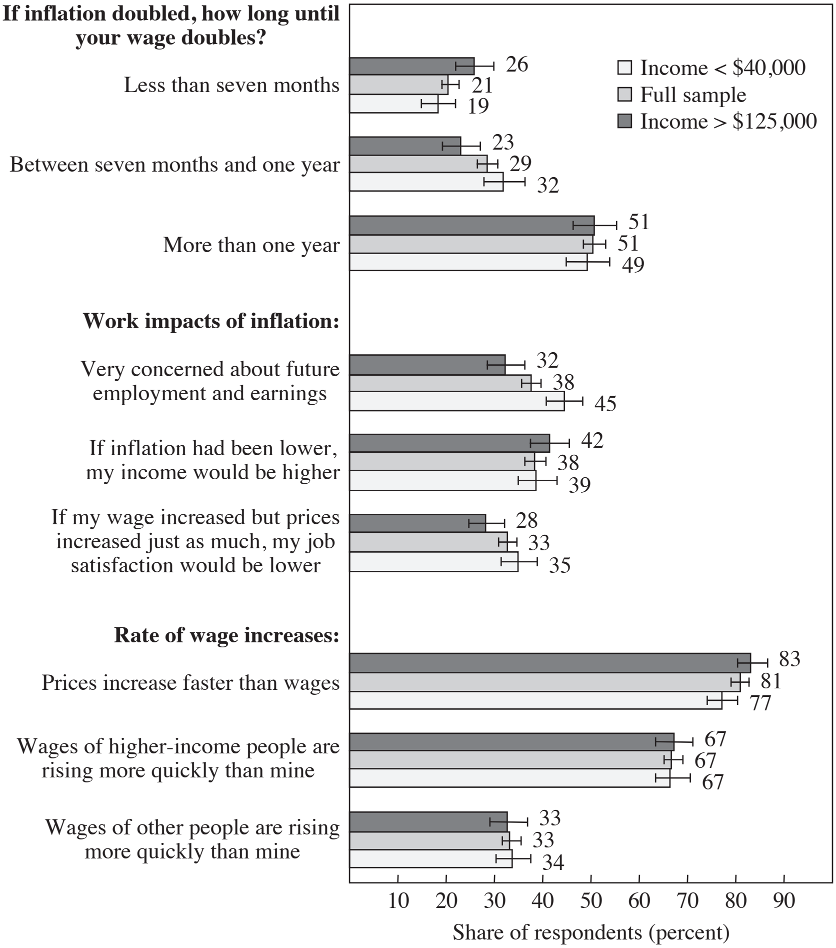
Figure 12. The Most Important Factor for Income Changes in the Past Five Years Has Been . . . [Open-Ended Text]



Source: Author's surveys.

Note: The figure reports the share of respondents whose answers belong to each category with 90 percent confidence intervals. The question is, "Think about how much your income (measured in dollars per month) went up (or down) in the past five years. What do you think are the most important factors that account for the change in your income? (Please try to list all the relevant factors that apply to you)." For each category, I report two example answers in online appendix A.3.6.

Figure 13. Inflation Impacts as a Worker



Source: Author's surveys.

Note: The figure reports the share of respondents whose answers are reflected by the statements listed alongside 90 percent confidence intervals. For more details on the questionnaire, see online appendix A.4.

of high-income respondents believe it will take less than seven months, they are nevertheless significantly more likely to do so than low-income respondents. Strikingly, these numbers are much lower than those in Shiller (1997) for the 1990s, when more than 80 percent of respondents thought it would take "several years" for their wage to adjust or that it would "never" adjust. Clearly, people have different perceptions of the labor market conditions today relative to that earlier time.

Furthermore, the share concerned about their future employment and earnings ranges from 32 percent among high-income respondents to 45 percent among low-income ones. Around 40 percent of respondents think that if inflation had been lower, their (nominal) income would be higher. In addition, one-third of respondents say that their job satisfaction would be *lower* if their wage increased just as much as prices. This share is quite similar to the one in Shiller (1997).

People systematically think that prices rise faster than wages (81 percent of all respondents).[8] Interestingly, two-thirds of respondents, including higher-income respondents, believe that the wages of higher-income people rise more quickly than theirs while only one-third believe that in general the wages of other people rise more quickly in response to inflation. There is therefore a clear sense of inequity in light of the wage adjustments to inflation.[9]

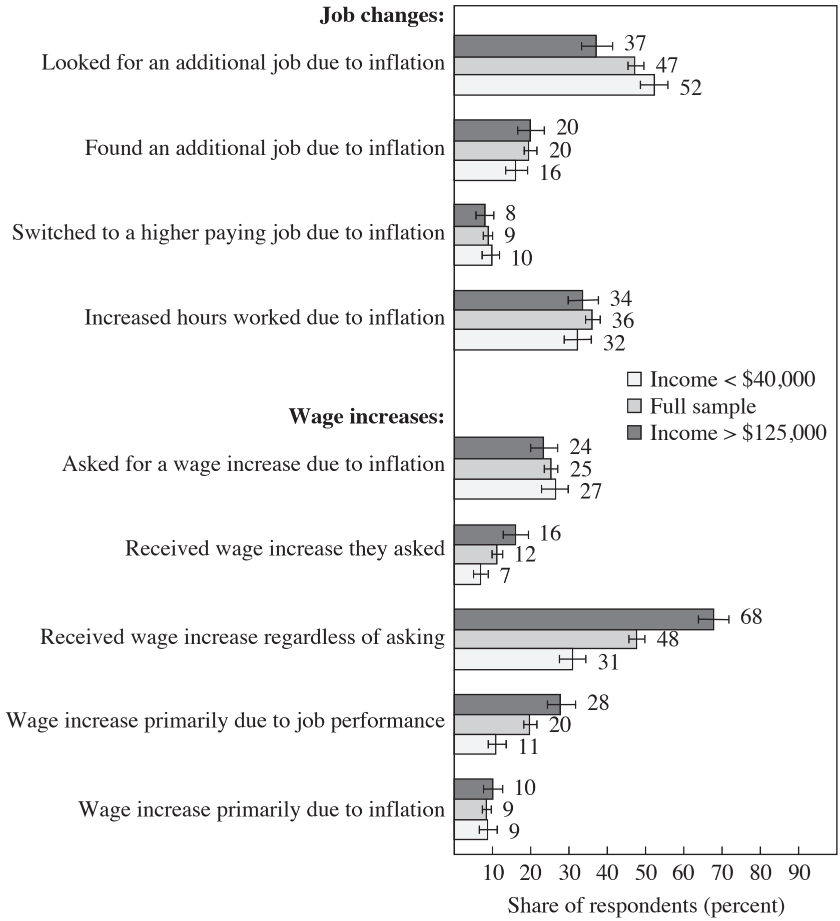
REACTIONS Faced with inflation, respondents appear to take various actions in the labor market (figure 14). But overall, they react more in their roles as consumers than as workers. Just around half of low-income respondents and a bit more than one-third of high-income respondents tried to look for an additional job (including part-time or gig work) because of inflation, but less than one-fifth report finding such a job. Less than 10 percent managed to switch to a higher paying job altogether because of inflation. Around one-third of people report trying to increase their on-the-job hours for extra income.[10] Respondents seem relatively reluctant to ask for wage increases because of inflation, with only one-quarter reporting having done so and about half of these reporting having received it. These results are in line with those in Pilossoph and Ryngaert (2023) and Hajdini and others (2022), who find that workers are relatively unlikely to search for a new job because of inflation, but the likelihood is higher among those with higher inflation expectations.

8. Data from the Bureau of Labor Statistics (available at https://www.bls.gov/charts/usual-weekly-earnings/usual-weekly-earnings-over-time-total-men-women.htm and from FRED at https://fred.stlouisfed.org/series/LES1252881600Q) indicate that the median usual weekly earnings of full-time wage and salary workers, quarterly averages, seasonally adjusted, evolved as follows since the start of the pandemic: 2019:Q4 +1.97% (relative to the previous year's Q4); 2020:Q4 +3.87%; 2021:Q4 –3.72%; 2022:Q4 +0.28%; 2023:Q4 +2.20%. As already discussed, these averages do not capture the inflation inequality across sectors, income groups, and places in the United States.

9. Sintos (2023) performs a comprehensive meta-analysis that shows that studies find, on average, small positive effects of inflation on inequality.

10. The data do not suggest that hours of work on average have increased over the last year; see Bureau of Labor Statistics, Average Weekly Hours of All Employees, Total Private [AWHAETP], accessed at https://fred.stlouisfed.org/series/AWHAETP.

Figure 14. Personal Reactions to Inflation as a Worker

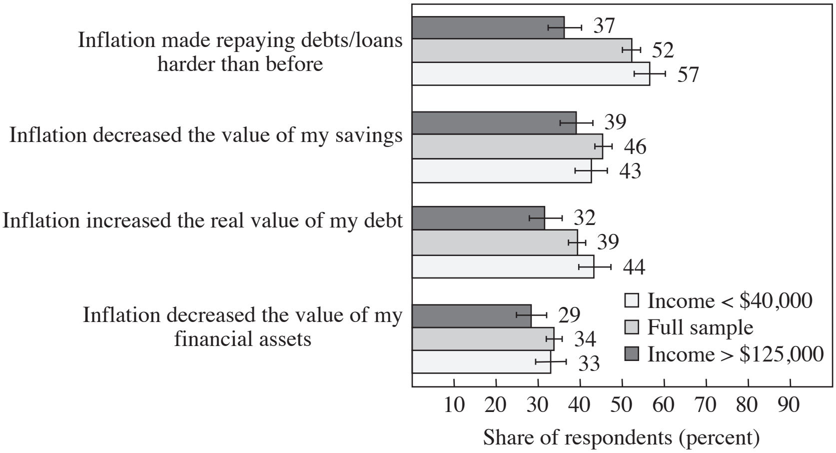


Source: Author's surveys.

Note: The figure reports the share of respondents whose answers are reflected by the statements listed alongside 90 percent confidence intervals. Note that all shares reported here are unconditional (e.g., 12 percent of the whole sample received the wage increase they asked for, not conditional on having asked for one). For more details on the questionnaire, see online appendix A.4.

Interestingly, people do not easily attribute wage increases to inflation. When it comes to *any* wage increase received (asked for or not), which happens to 48 percent of respondents, more respondents (20 percent) will attribute the raise primarily to their on-the-job performance than primarily to inflation (9 percent), with the remaining share attributing it to a mix of the two. That discrepancy is particularly pronounced among high-income

Figure 15. Inflation Impacts as an Asset Holder



Source: Author's surveys.
Note: The figure reports the share of respondents whose answers are reflected by the statements listed alongside 90 percent confidence intervals. For more details on the questionnaire, see online appendix A.4.

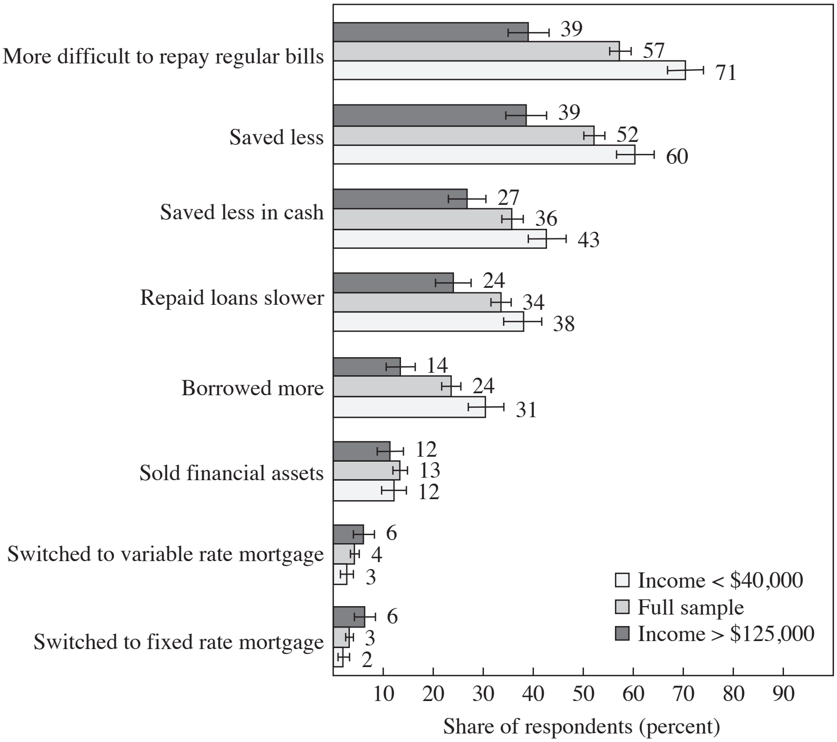
respondents, where 28 percent attribute it to performance primarily, and 10 percent to inflation only. In the online appendix, table A1 shows that when the wage increase occurs during a job change, respondents are more likely to attribute it to on-the-job performance and career progression than if it happens in the same job. Therefore, it seems that people are reluctant to perceive wage increases as the result of inflation adjustments rather than performance.

III.C. As an Asset Holder

IMPACTS Inflation can also have an impact on people who have assets or liabilities. Figure 15 shows that, among low-income respondents, 57 percent believe that inflation has made repaying their debt or loans harder, 44 percent think it has increased the *real* value of their debt (which we explicitly define as "the amount you owe in relation to the general cost of living and prices"), and 43 percent believe it has decreased the value of their savings. These shares are consistently lower among high-income respondents.

REACTIONS Respondents, especially low-income ones, also react along the savings and borrowing margins in response to inflation (figure 16). Seventy-one percent among low-income respondents have more difficulty paying their regular bills and, as a result, save less (60 percent), repay their

Figure 16. Personal Reactions to Inflation as an Asset Holder



Source: Author's surveys.
Note: The figure reports the share of respondents whose answers are reflected by the statement listed alongside 90 percent confidence intervals. For more details on the questionnaire, see online appendix A.4.

loans more slowly (38 percent), and borrow more (31 percent). Higher-income respondents also report these behaviors, but to a much lesser degree.

Interestingly, only around 36 percent of all respondents shift the composition of their savings away from cash in response to inflation (the question explicitly asked about the composition, rather than the total amount of savings, which, as just discussed, also declines). A very small share of respondents (between 3 and 4 percent) switch their type of mortgage from variable rate to fixed rate or vice versa.

III.D. Psychological and Emotional Impacts of Inflation

Given all these perceived impacts of inflation on people, as consumers, workers, and asset holders, one can reasonably expect that there would be psychological and emotional impacts too.

EMOTIONS Figure 17 plots an emotion analysis, performed using the RoBERTa model to classify answers to the open-ended question, "What feelings do you typically experience when you hear news reports about 'rising inflation'?"[11] A first interesting finding is that around 40 percent of respondents do not report specific emotions in response to that sentence. However, that share is only 31 percent among low-income respondents compared to 50 percent among high-income ones. Low-income respondents are much more likely to report despair, stress, or fear. Reported emotions are relatively balanced by political leaning.

WHO ARE YOU ANGRY AT? I also asked a question that mimics one in Shiller (1997) and is specifically about anger in a concrete context (rather than just abstractly thinking about inflation news). The question reads, "When you went to the store and saw that prices were higher, did you feel a little angry?"[12] In this more specific context, 43 percent of respondents answer "Yes, often," 44 percent answer "Yes, sometimes," and 13 percent answer "No, never." These numbers are very close to the ones in Shiller (1997) (38 percent, 48 percent, and 15 percent, respectively).

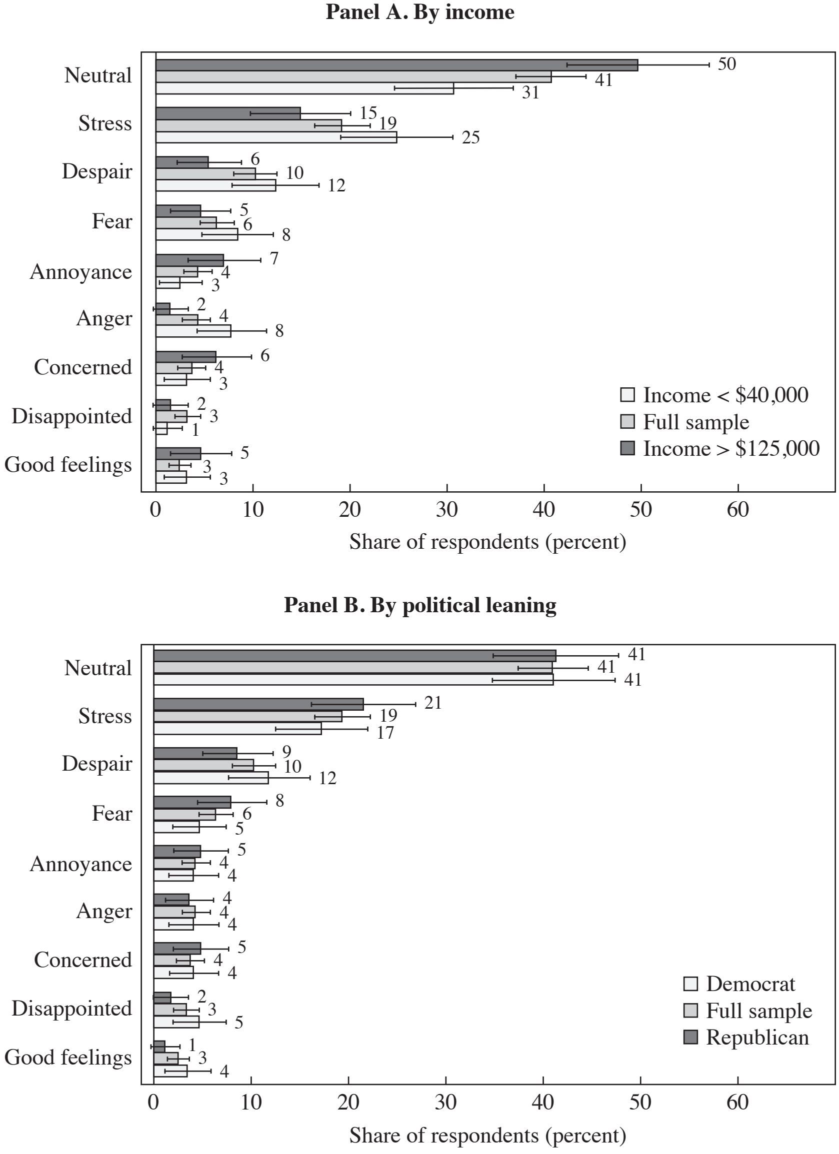
As a follow-up open-ended question, respondents who answered that they are at least somewhat angry were asked, "Who do you tend to feel angry at?" Figure 18 plots the distribution of answers, which can be classified into four major categories: the government overall, mentioned by 37 percent of all respondents ("I'm angry because the price rise could have been prevented. Instead, it was allowed to happen by the government. I do not blame the business owners though because it was forced upon them," "The government claiming that it is working for the middle-class Americans, while simultaneously destroying it"), although there is a smaller but sizable group of people who explicitly focus on Biden ("Joe Biden, for trying to use helicopter money to buy votes"). As might be expected given the current political leaning of the government, it is especially Republican respondents who blame the government or Biden.

The second most mentioned category is businesses ("The big corporations that won't let their profits fall by even one percent and give the

11. The model is publicly available at https://huggingface.co/SamLowe/roberta-base-go_emotions. It is a 125,000-parameter RoBERTa-base model trained on the GoEmotions data set for multilabel classification. It has twenty-eight possible emotions, and for each input the model assigns a probability distribution over these labels. As is standard in the literature, I tag each answer with the emotion classified with the highest probability, as long as the probability is greater than 0.5. Otherwise, I leave it nonlabeled.

12. The question in Shiller adds "at someone" at the end of the question, namely, "When you go to the store and see that prices are higher, do you sometimes feel a little angry at someone?" I thought it is not necessary to prime people about being angry at someone.

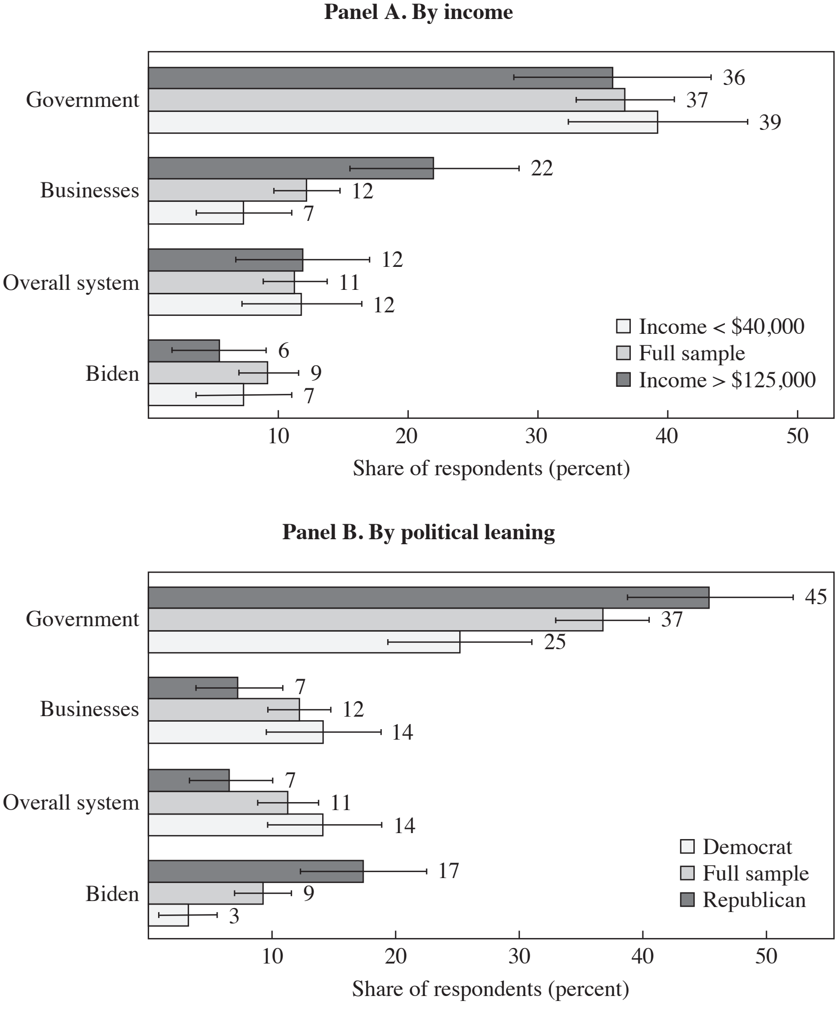
Figure 17. When Hearing Rising Inflation I Feel . . . [Open-Ended Text]



Source: Author's surveys.

Note: The figure reports the share of respondents whose answers belong to each category with 90 percent confidence intervals. The precise question is, "What feelings do you typically experience when you hear news reports about 'rising inflation'?" The categorization was carried out by the RoBERTa emotion model. I only report emotions mentioned by at least ten respondents. I assign to each respondent their most likely emotion and do not assign any emotion if all probabilities are lower than 0.5. For each category, I report some keywords in online appendix A.3.7.

Figure 18. When I Went to the Store and Saw That Prices Were Higher, I Felt Angry at . . . [Open-Ended Text]



Source: Author's surveys.

Note: The figure reports the share of respondents whose answers are reflected by the statement listed alongside 90 percent confidence intervals. All the shares reported here are unconditional. For the categories "Government," "Businesses," "Biden," and "Overall system," I report three example answers in online appendix A.3.8. For more details on the questionnaire, see online appendix A.5.

customer the tax at the end when they should be paying the tax," "The people causing inflation and the corporations who aren't willing to lose any profit growth," and "The corporations who have to keep up their huge bonuses to their top people"). This is especially the case among Democrats and, interestingly, high-income respondents. Finally, people also mention the system overall ("Not so much angry at a specific person just the overall situation because people like me who are on a budget now have to learn to make that budget stretch thinner than we were already" and "The entire system").

STRESS CAUSED BY INFLATION To probe further into the psychological impacts of inflation, I present respondents with a series of closed-ended, more specific questions. Figure 19 shows that 70 percent of respondents would be less stressed if inflation had been lower and three-quarters believe that inflation has worsened their outlook on their future economic well-being. Stress seems to have affected all income groups, but for different reasons. The lower bars of the figure show that among lower-income respondents, stress is mainly due to the inability to afford essentials (for 44 percent of respondents who report feeling more stressed) and the inability to pay rent (among 24 percent of them). For higher-income respondents, stress is caused by investment losses (37 percent of respondents) and, to a lesser extent, cutting down on going out and holidays and paying their mortgage or college tuition for their children.[13]

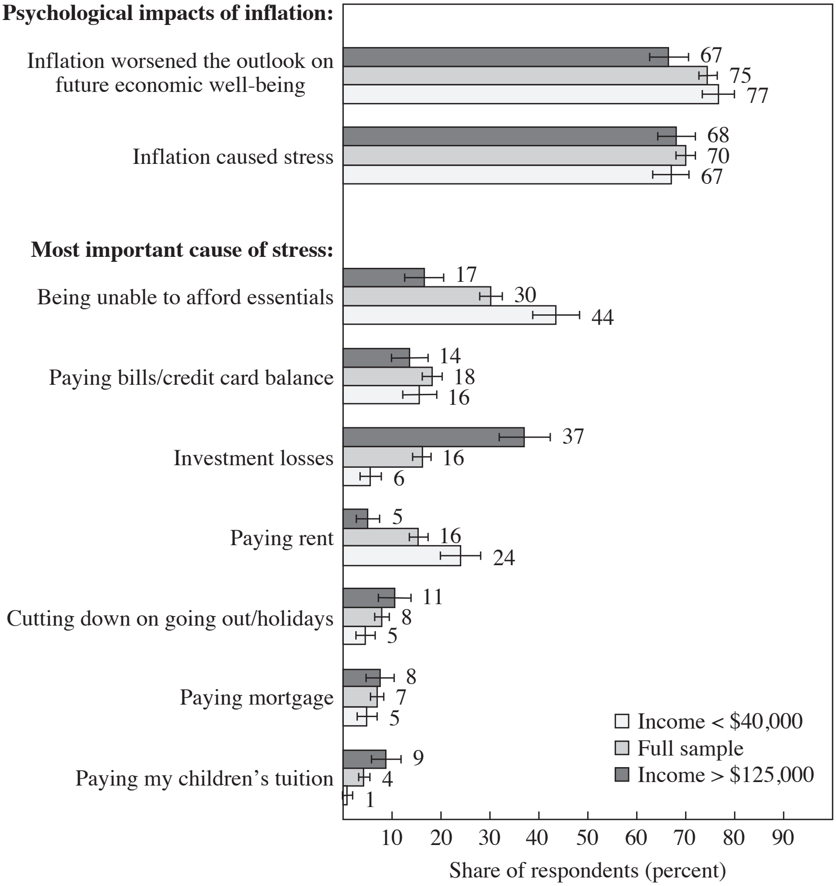
IV. Policy Views

IV.A. Priority of Inflation

Given the personal impacts and costs of inflation, one might expect inflation to rank high in respondents' political priorities. Therefore, I ask respondents to rank various economic and social issues, including inflation. The top bars in figure 20 report the share of respondents who rank a given economic issue first. The bottom set of bars shows the ranking among social issues. Among both sets of issues, inflation most often ranks first, much more so among social than economic issues. About one-third of respondents rank it first among economic issues, ahead of financial stability, economic growth, low unemployment, and national defense; 41 percent rank it first among social issues, ahead of health care, civil rights, education, gun rights, and abortion. There are interesting political gaps along the

13. All these shares are conditional on reporting that inflation caused stress.

Figure 19. Inflation's Psychological Impacts



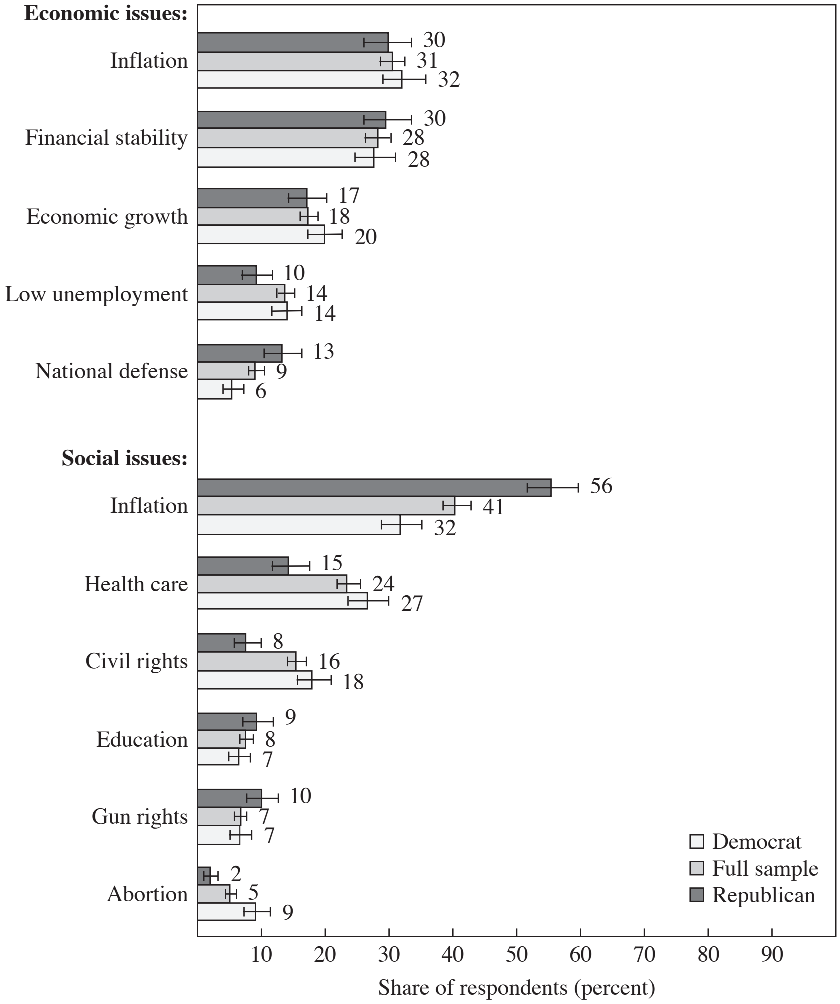
Source: Author's surveys.
Note: The figure reports the share of respondents whose answers are reflected by the statement listed alongside 90 percent confidence intervals. The shares shown for the second set of bars ("Most important cause of stress") are conditional on reporting that inflation caused stress. For more details on the questionnaire, see online appendix A.4.

social issue dimension, with Republicans much more likely to rank inflation higher up, while Democrats are almost tied between inflation and health care. But there is bipartisan agreement on the ranking of economic issues.

IV.B. The Inflation-Unemployment Trade-Off

A salient trade-off for economists under some circumstances is that between inflation and unemployment. How do respondents perceive this

Figure 20. Ranking of Social and Economic Issues



Source: Author's surveys.

Note: The figure reports the share of respondents choosing the listed statement as the most important one alongside 90 percent confidence intervals. For more details on the questionnaire, see online appendix A.4.

trade-off? An overwhelming majority of respondents believe that inflation and unemployment are related. However, only one-quarter believe that they are negatively related. Clearly, people associate high inflation with economic downturns and higher unemployment, a view consistent with stagflation. Indeed, figure 21 also shows that 70 percent of all respondents believe that "inflation indicates a poor state of the economy." Relatedly, a majority of respondents, especially among Republicans, also believe that inflation decreases exports.

These results echo those in Shiller (1997), where few respondents thought that low unemployment was a potential benefit of inflation. It also resonates with the open-ended question studied above, where almost no respondents were able to think of potential upsides to inflation.

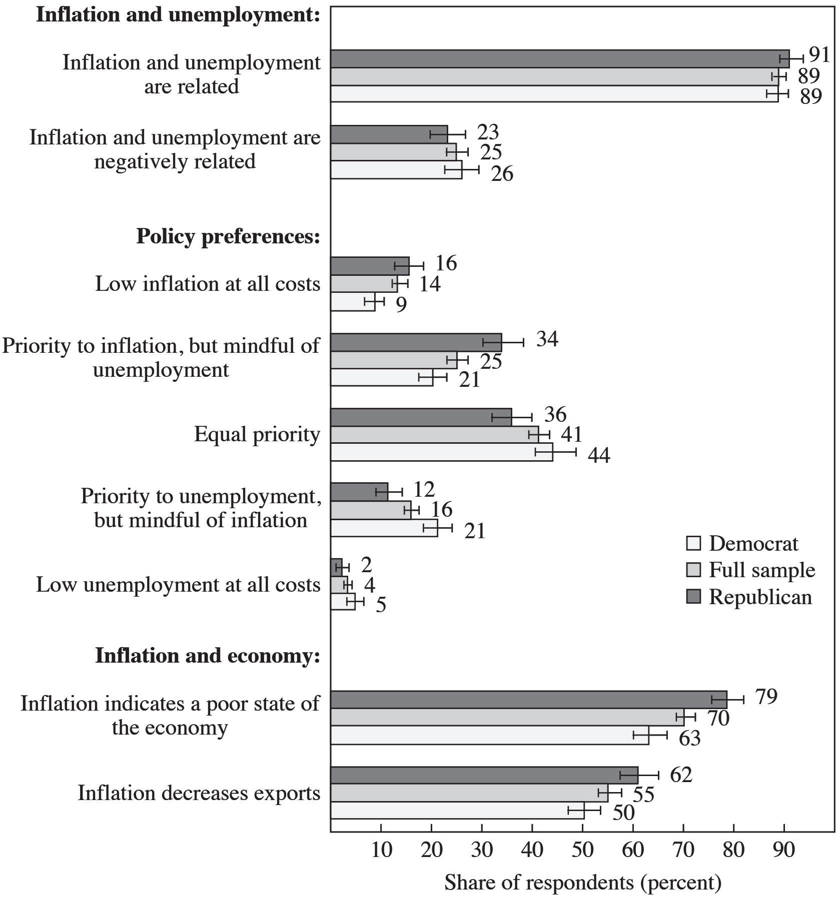
If I ask respondents to express their preferences between low inflation and low unemployment in a very simple way, 41 percent select "equal priority" and one-quarter select "priority to inflation, but mindful of unemployment," consistent with the rankings observed above. Republican respondents put significantly higher weight on low inflation relative to low unemployment, while Democrats are more evenly divided (see figure 21). Online appendix figure A13 shows that lower-income respondents are more likely to put equal priority on inflation and unemployment, while higher-income ones slightly emphasize low inflation.[14]

V. Conclusion

Insights from two new surveys on inflation discussed in this paper reveal people's aversion to inflation, which is deeply rooted in its perceived impact on their financial well-being and the broader economy. The main concern highlighted is the erosion of purchasing power, with many feeling that wage growth does not keep up with the pace of rising prices. This situation leads to significant reported adjustments in spending habits, particularly among lower-income individuals, who often find themselves postponing or reducing the quality and quantity of their purchases. The study also points to a widespread perception of inequality exacerbated by inflation, as respondents believe that high-income earners' wages increase more rapidly in inflationary periods, further deepening the divide between different income groups.

14. The perceived and desired trade-offs between inflation and unemployment are studied in Binetti, Nuzzi, and Stantcheva (2024).

Figure 21. The Perceived Inflation versus Unemployment Trade-Off



Source: Author's surveys.
Note: The figure reports the share of respondents whose answers are reflected by the statement listed alongside 90 percent confidence intervals. The share reporting those who say that inflation and unemployment are negatively related is conditional on saying they are related. For more details on the questionnaire, see online appendix A.4.

Responses to inflation also include stress and emotional reactions, reflecting another potential personal and societal toll of rising prices. There is a clear division in opinions on the causes of inflation, with political affiliations influencing whether individuals blame the government, businesses, or broader systemic factors. There is a consensus on the lack of positive outcomes from inflation, with few recognizing any positive associations or trade-offs, such as with lower unemployment or economic growth. Instead, inflation is predominantly associated with negative economic and social effects, making it a high priority for policy action. This aligns with the earlier findings from the 1990s by Shiller (1997).

The perceived unequal consequences of inflation by income groups are in line with recent empirical evidence on the heterogeneous impacts of inflation. It would be valuable to dig deeper into people's understanding of inflation, in terms of its causes and consequences and how it relates to other economic outcomes, as well as to understand what drives their views on how policy should address this.

ACKNOWLEDGMENTS I thank Carola Binder, Janice Eberly, Yuriy Gorodnichenko, Francesco Nuzzi, and Jón Steinsson for helpful comments and feedback. I am deeply grateful to Alberto Binetti, Filippo Giorgis, and Alfonso Merendino for excellent research assistance.

References

Aklin, Michaël, Eric Arias, and Julia Gray. 2022. “Inflation Concerns and Mass Preferences over Exchange-Rate Policy.” *Economics and Politics* 34, no. 1: 5–40.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. 2020. “The Polarization of Reality.” *American Economic Association Papers and Proceedings* 110: 324–28.

Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart. 2021. “Narratives about the Macroeconomy.” ECONtribute Discussion Paper 127. Bonn: University of Bonn and University of Cologne.

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. 2022. “Subjective Models of the Macroeconomy: Evidence from Experts and Representative Samples.” *Review of Economic Studies* 89, no. 6: 2958–91.

Angelico, Cristina, and Federica Di Giacomo. 2019. “Heterogeneity in Inflation Expectations and Personal Experience.” Working Paper. https://events.ceu.edu/sites/default/files/media/attachment/jmp_angelico_digiacomo_2019.pdf.

Argente, David, and Munseob Lee. 2021. “Cost of Living Inequality during the Great Recession.” *Journal of the European Economic Association* 19, no. 2: 913–52.

Atkin, David, Benjamin Faber, Thibault Fally, and Marco Gonzalez-Navarro. 2024. “Measuring Welfare and Inequality with Incomplete Price Information.” *Quarterly Journal of Economics* 139, no. 1: 419–75.

Bachmann, Rüdiger, Tim O. Berg, and Eric R. Sims. 2015. “Inflation Expectations and Readiness to Spend: Cross-Sectional Evidence.” *American Economic Journal: Economic Policy* 7, no. 1: 1–35.

Binder, Carola, Wesley Janson, and Randal Verbrugge. 2023. “Out of Bounds: Do SPF Respondents Have Anchored Inflation Expectations?” *Journal of Money, Credit and Banking* 55, no. 2–3: 559–76.

Binder, Carola, Tucker S. McElroy, and Xuguang S. Sheng. 2022. “The Term Structure of Uncertainty: New Evidence from Survey Expectations.” *Journal of Money, Credit and Banking* 54, no. 1: 39–71.

Binetti, Alberto, Francesco Nuzzi, and Stefanie Stantcheva. 2024. “People’s Understanding of Inflation.” Working Paper 32497. Cambridge, Mass.: National Bureau of Economic Research.

Bolhuis, Marijn A., Judd N. L. Cramer, Karl Oskar Schulz, and Lawrence H. Summers. 2024. “The Cost of Money Is Part of the Cost of Living: New Evidence on the Consumer Sentiment Anomaly.” Working Paper 32163. Cambridge, Mass.: National Bureau of Economic Research.

Bruine de Bruin, Wändi, Wilbert van der Klaauw, Julie S. Downs, Baruch Fischhoff, Giorgio Topa, and Olivier Armantier. 2010. “Expectations of Inflation: The Role of Demographic Variables, Expectation Formation, and Financial Literacy.” *Journal of Consumer Affairs* 44, no. 2: 381–402.

Bruine de Bruin, Wändi, Wilbert van der Klaauw, and Giorgio Topa. 2011. “Expectations of Inflation: The Biasing Effect of Thoughts about Specific Prices.” *Journal of Economic Psychology* 32, no. 5: 834–45.

Cavallo, Alberto. 2024. “Inflation with COVID Consumption Baskets.” *IMF Economic Review* 72, no. 2: 902–17. https://www.hbs.edu/faculty/Pages/item.aspx?num=64618.

Cavallo, Alberto, Guillermo Cruces, and Ricardo Perez-Truglia. 2017. “Inflation Expectations, Learning, and Supermarket Prices: Evidence from Survey Experiments.” *American Economic Journal: Macroeconomics* 9, no. 3: 1–35.

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, and Michael Weber. 2023. “Forward Guidance and Household Expectations.” *Journal of the European Economic Association* 21, no. 5: 2131–71.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. “Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts.” *American Economic Review* 105, no. 8: 2644–78.

Coibion, Olivier, Yuriy Gorodnichenko, and Rupal Kamdar. 2018. “The Formation of Expectations, Inflation, and the Phillips Curve.” *Journal of Economic Literature* 56, no. 4: 1447–91.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2022. “Monetary Policy Communications and Their Effects on Household Inflation Expectations.” *Journal of Political Economy* 130, no. 6: 1537–84.

Coles, Jeffrey L., and Paul Chen. 1990. “Preferences for Unemployment versus Inflation.” *Applied Economics* 22, no. 3: 347–58.

D’Acunto, Francesco, Daniel Hoang, Maritta Paloviita, and Michael Weber. 2019. “Cognitive Abilities and Inflation Expectations.” *American Economic Association Papers and Proceedings* 109: 562–66.

D’Acunto, Francesco, Ulrike Malmendier, Juan Ospina, and Michael Weber. 2021. “Exposure to Grocery Prices and Inflation Expectations.” *Journal of Political Economy* 129, no. 5: 1615–39.

Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva. 2022. “Fighting Climate Change: International Attitudes toward Climate Policies.” Working Paper 30265. Cambridge, Mass.: National Bureau of Economic Research.

Di Tella, Rafael, Robert J. MacCulloch, and Andrew J. Oswald. 2001. “Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness.” *American Economic Review* 91, no. 1: 335–41.

Easterly, William, and Stanley Fischer. 2001. “Inflation and the Poor.” *Journal of Money, Credit and Banking* 33, no. 2: 160–78.

Ferrario, Beatrice, and Stefanie Stantcheva. 2022. “Eliciting People’s First-Order Concerns: Text Analysis of Open-Ended Survey Questions.” *American Economic Association Papers and Proceedings* 112: 163–69.

Goldfayn-Frank, Olga, and Johannes Wohlfart. 2020. “Expectation Formation in a New Environment: Evidence from the German Reunification.” *Journal of Monetary Economics* 115: 301–20.

Hajdini, Ina, Edward S. Knotek II, John Leer, Mathieu Pedemonte, Robert W. Rich, and Raphael Schoenle. 2022. “Low Passthrough from Inflation Expectations to Income Growth Expectations: Why People Dislike Inflation.” Working

Paper 22-21R. Federal Reserve Bank of Cleveland. https://www.clevelandfed.org/publications/working-paper/2022/wp-2221-low-passthrough-from-inflation-expectations-to-income.

Hofstetter, Marc, and José Nicolás Rosas. 2021. "The Poor and the Rich: Preferences over Inflation and Unemployment." *Journal of Money, Credit and Banking* 53, no. 4: 875–95.

Howarth, David, and Charlotte Rommerskirchen. 2017. "Inflation Aversion in the European Union: Exploring the Myth of a North–South Divide." *Socio-Economic Review* 15, no. 2: 385–404.

Hübner, Malte, and Marcus Klemm. 2015. "Preferences over Inflation and Unemployment in Europe: A North–South Divide?" *International Review of Economics* 62: 319–35.

Jain, Monica, Olena Kostyshyna, and Xu Zhang. 2022. "How Do People View Price and Wage Inflation?" Staff Working Paper 2022-34. Ottawa: Bank of Canada.

Jaravel, Xavier. 2021. "Inflation Inequality: Measurement, Causes, and Policy Implications." *Annual Review of Economics* 13, no. 1: 599–629.

Jaravel, Xavier, and Alan Olivi. 2021. "Prices, Non-homotheticities, and Optimal Taxation: The Amplification Channel of Redistribution." Working Paper. Social Science Research Network. http://dx.doi.org/10.2139/ssrn.3383693.

Jayadev, Arjun. 2008. "The Class Content of Preferences towards Anti-inflation and Anti-unemployment Policies." *International Review of Applied Economics* 22, no. 2: 161–72.

Kamdar, Rupal. 2019. "The Inattentive Consumer: Sentiment and Expectations." Working Paper. https://rupalkamdar.github.io/pdfs/Inattentive_Consumer.pdf.

Kaplan, Greg, and Sam Schulhofer-Wohl. 2017. "Inflation at the Household Level." *Journal of Monetary Economics* 91: 19–38.

Kim, Gwangmin, and Carola Binder. 2023. "Learning-through-Survey in Inflation Expectations." *American Economic Journal: Macroeconomics* 15, no. 2: 254–78.

Malmendier, Ulrike, and Stefan Nagel. 2016. "Learning from Inflation Experiences." *Quarterly Journal of Economics* 131, no. 1: 53–87.

Pilossoph, Laura, and Jane M. Ryngaert. 2023. "Job Search, Wages, and Inflation." Working Paper. https://wustl.app.box.com/s/u61tn6wi938402ix6k4m8zus451qn3ks.

Rotemberg, Julio J. 2005. "Customer Anger at Price Increases, Changes in the Frequency of Price Adjustment and Monetary Policy." *Journal of Monetary Economics* 52, no. 4: 829–52.

Rotemberg, Julio J. 2011. "Fair Pricing." *Journal of the European Economic Association* 9, no. 5: 952–81.

Ruprah, Inder J., and Pavel Luengas. 2011. "Monetary Policy and Happiness: Preferences over Inflation and Unemployment in Latin America." *Journal of Socio-Economics* 40, no. 1: 59–66.

Scheve, Kenneth. 2003. "Public Demand for Low Inflation." Working Paper 172. London: Bank of England.

Scheve, Kenneth. 2004. "Public Inflation Aversion and the Political Economy of Macroeconomic Policymaking." *International Organization* 58, no. 1: 1–34.

Shiller, Robert J. 1997. "Why Do People Dislike Inflation?" In *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer. Chicago: University of Chicago Press.

Sintos, Andreas. 2023. "Does Inflation Worsen Income Inequality? A Meta-analysis." *Economic Systems* 47, no. 4: 101146.

Stantcheva, Stefanie. 2021. "Understanding Tax Policy: How Do People Reason?" *Quarterly Journal of Economics* 136, no. 4: 2309–69.

Stantcheva, Stefanie. 2023a. "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible." *Annual Review of Economics* 15: 205–34.

Stantcheva, Stefanie. 2023b. "Understanding of Trade." Working Paper 30040. Cambridge, Mass.: National Bureau of Economic Research.

van Lelyveld, Iman. 1999. "Inflation or Unemployment? Who Cares?" *European Journal of Political Economy* 15, no. 3: 463–84.

Weber, Michael, Francesco D'Acunto, Yuriy Gorodnichenko, and Olivier Coibion. 2022. "The Subjective Inflation Expectations of Households and Firms: Measurement, Determinants, and Implications." *Journal of Economic Perspectives* 36, no. 3: 157–84.

Wimer, Christopher, Sophie Collyer, and Xavier Jaravel. 2019. "The Costs of Being Poor: Inflation Inequality Leads to Three Million More People in Poverty." Policy Brief. New York: Center on Poverty and Social Policy, Columbia University.

Comments and Discussion

COMMENT BY

CAROLA BINDER When Robert Shiller (1997) conducted his famous study of public attitudes toward inflation, countries around the world had only recently endured painful episodes of high unemployment and low output in order to reduce inflation from very high levels (Romer and Romer 1997). There was a consensus that this trade-off was necessary, but this consensus was difficult to reconcile with standard economic theory (Wen 2010). Economists modeled the welfare cost of inflation as coming from the tax it imposed on real money balances, measured as the area under the money demand function corresponding to the deadweight loss of moving from a lower to a higher inflation rate (Bailey 1956). By this measure, inflation had surprisingly small costs.

Thus, in their widely used textbook, Blanchard and Fischer noted that "standard characterizations of the policymaker's objective function put more weight on the costs of inflation than is suggested by our understanding of the effects of inflation; in doing so, they probably reflect political realities and the heavy political costs of high inflation" (1989, 567–68).

Shiller took what was, at the time, an unusual approach for an economist. He *asked people* about their beliefs and preferences. In doing so, he rejected Samuelson's (1938) revealed preference theory—"one of the most influential ideas in economics" (Varian 2006, 99)—as the only or best method of understanding consumer behavior. To suggest that consumers could simply tell economists their preferences was as unorthodox as more recent "neuroeconomics" research (of which Shiller is also a fan), which uses brain scans to study consumer behavior (Shiller 2011).

Brookings Papers on Economic Activity, Spring 2024: 47–65

Shiller found that people in the United States, Germany, and Brazil widely believed that inflation eroded their standard of living; they did not believe that their income kept up with rising prices. They believed that controlling inflation was one of the most important goals of economic policy. And while Shiller did not speak directly to the policymaker's objective function, he did find that people said they would prefer ten years of 2 percent annual inflation and 9 percent unemployment over ten years of 10 percent monthly inflation and 3 percent unemployment. This hypothetical trade-off was maybe too extreme to be useful; Christina Romer and David Romer, who edited the National Bureau of Economic Research (NBER) volume in which Shiller's work appeared, noted that "while there is ample evidence that high inflation harms economic growth and stability, there is remarkably little research on the costs and benefits of reducing inflation from, say, 3% to 1%" (1997, 1). It is not really clear what, if anything, Shiller's results imply about those costs or benefits, and in Mankiw's discussion of Shiller's results, he said that "I am not at all sure in what direction they should push either economic theory or economic policy" (1997, 65).

A few decades and one high inflation episode later, Stantcheva finds similar results for US consumers. People still dislike inflation, believe that it erodes their purchasing power, and rank it as one of our country's biggest problems. Like Shiller, she avoids making explicit policy recommendations based on these results, but surely, questions about the implications for policymakers' objective functions will be at the front of mind for any reader. Does consumers' reported distaste for inflation justify putting more weight on inflation in the objective function or perhaps lowering the inflation target? In the next recession, should policymakers be more cautious in their fiscal and monetary response?

DO PEOPLE DISLIKE INFLATION? To start, let us consider what happened in between Shiller's and Stantcheva's surveys. In particular, I want to reflect on attitudes toward inflation in the years following the Great Recession. When the Federal Open Market Committee (FOMC) announced its 2 percent inflation target in January 2012, with the unemployment rate at 8.3 percent, they promised to follow "a balanced approach" in promoting price stability and maximum employment (Federal Reserve Board of Governors 2012). Unemployment fell very gradually, reaching 5 percent in December 2015.[1] Although the Personal Consumption Expenditures (PCE) inflation was still well below target, at around 1.1 percent, the FOMC raised rates

1. Bureau of Labor Statistics, series UNRATE, retrieved from FRED, https://fred.stlouisfed.org/series/UNRATE.

for the first time since the recession, in anticipation that inflation would soon begin to rise (Federal Reserve Board of Governors 2015).[2] The Fed's focus on price stability was widely criticized, especially by progressive groups representing labor and consumer interests (Binder 2024). The Fed Up coalition, made up of left-leaning and populist advocacy groups, community organizations, and labor unions, urged against additional rate hikes on the grounds that the benefits of full employment far outweighed the costs of a little inflation.[3]

This sentiment became quite influential and was repeated at the *Fed Listens* events conducted in 2019 as part of the Fed's framework review. The *Fed Listens* report notes that "there was less discussion at the *Fed Listens* events of inflation than there was of labor market conditions" and that "during the roundtable discussion, one participant argued that some inflation is good and echoed a sentiment from the advisory group discussions—that today inflation may be too low" (Federal Reserve Board of Governors 2020a, 7, 46). It also notes that "younger participants noted that their generation is more concerned with another recession than with high inflation" (ibid., 46).

Following this listening campaign, the Fed amended its framework by adopting average inflation targeting in 2020. The revised Statement on Longer-Run Goals and Monetary Policy Strategy explains that "following periods when inflation has been running persistently below 2 percent, appropriate monetary policy will likely aim to achieve inflation moderately above 2 percent for some time" (Federal Reserve Board of Governors 2020b, par. 4). The new framework is deliberately asymmetric, promising to make up for inflation undershoots but not overshoots. With the new framework, the Fed indicated that it would not do what it did in 2015: it would not raise rates preemptively in anticipation of inflation but instead would wait for inflation to actually appear. As a result, the Fed delayed tightening policy in 2021 (Eggertsson and Kohn 2023).

In other words, the Fed listened when people said that they didn't mind inflation so much. And this wasn't the first time. Our monetary institutions owe a lot to people's dislike of *deflation*. Falling prices, which increased farmers' real debt burdens, were extremely unpopular in our country's early years. The gold standard, which limited the possibility of major inflation,

2. Bureau of Economic Analysis, series PCEPILFE, retrieved from FRED, https://fred.stlouisfed.org/series/PCEPILFE.

3. Center for Popular Democracy, "Building a National Campaign for a Strong Economy: Fed Up," https://www.populardemocracy.org/campaign/building-national-campaign-strong-economy-fed.

also sometimes brought about episodes of deflation. By the time of William Jennings Bryan, populist politicians were the biggest advocates of leaving the gold standard and enabling a more expansionary monetary policy that they thought the people would prefer (Binder 2024). We eventually learned that without an independent central bank, politicians are tempted to create an excessive amount of inflation in the hopes of pleasing the people.

Our monetary institutions are deliberately designed to give policymakers the power and discretion to create inflation if they choose, with some safeguards against the longer-run consequences that would come from succumbing to our short-run taste for monetary expansion. The idea of constraining policymakers even more tightly in the interest of preventing inflation altogether is very unpopular. Why, then, do people report that they dislike inflation, and what should we make of these survey results?

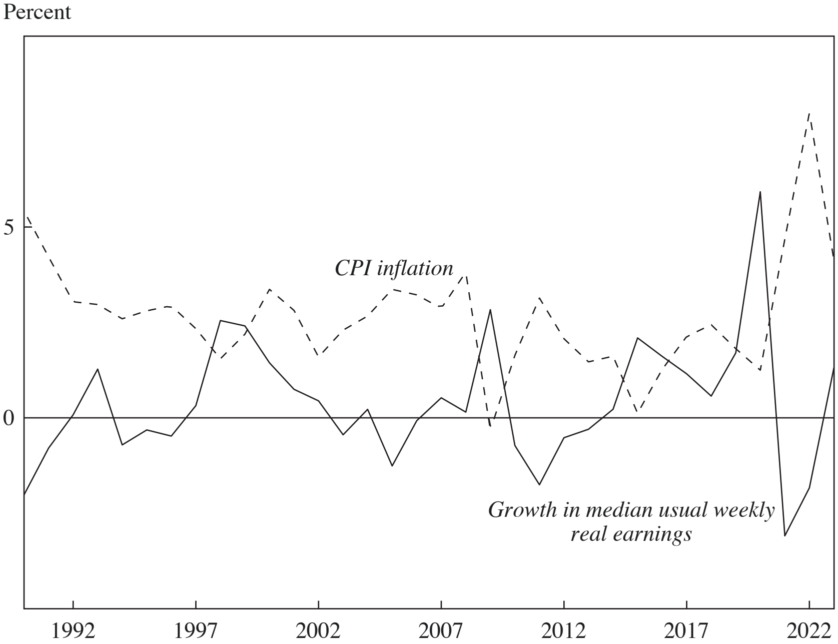
INTERPRETING THE SURVEYS Shiller (1997) surveyed not only consumers but also economists about inflation, and he found that economists and noneconomists viewed inflation very differently. In Mankiw's discussion of Shiller's paper, he noted that the principal finding was an "inflation fallacy." Laymen, unlike economists, "say that inflation makes them poorer. . . . It is tempting for economists to snicker at this answer. Such a reaction gives us a sense of superiority, and it offers an opportunity to reciprocate the low regard in which much of the public holds the economics profession" (1997, 66).

Unfortunately, Stantcheva did not send her survey to economists, so we cannot compare economists' and laymen's interpretations of recent inflation. But we should still resist the temptation to snicker at their answers.

First, inflation can be associated with lower real wages and living standards, particularly if it is supply-driven. Mankiw suggested that you could get at this idea by phrasing a question such as: "A shock hits the economy. One result of the shock is a higher cost of living, as measured by the consumer price index. What is the likely effect of this shock on your standard of living?" (1997, 66). Mankiw regressed annual nominal GDP growth on annual GDP deflator inflation from 1959 to 1994 and found a coefficient around 0.6 (standard error 0.14). He concluded that "when inflation is high, growth in nominal income is also high, but not by enough to compensate fully for the change in prices. Shocks to aggregate supply seem a natural explanation for this result" (1997, 66).

In more recent years, the coefficient is above one, though one is in the 95 percent confidence interval. From 2004 through 2023, for example, the coefficient is 1.5, with a standard error of 0.41. In Mankiw's interpretation, then, monetary shocks have caused real output and inflation to move in the same direction, and supply shocks are less dominant. But supply shocks

Figure 1. Inflation and Real Wage Growth Are Negatively Correlated



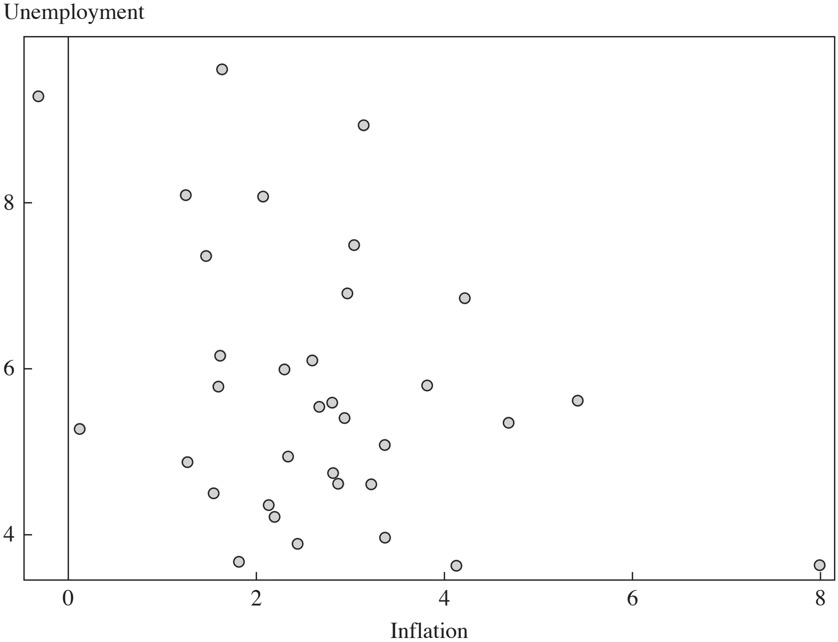
Source: Bureau of Labor Statistics, series CPIAUCSL and LES1252881600Q, retrieved from FRED, Federal Reserve Bank of St. Louis.

Note: Figure shows Consumer Price Index (CPI) inflation and growth in median usual weekly real earnings of wage and salary workers 16 years and older. Both series are annual, and the percent change from the previous year is shown. Correlation between the two series is –0.61.

are still a possible driver of inflation, and it is not crazy for consumers to recognize that some types of inflation are associated with lower real wages. In fact, inflation and real wage growth are strongly negatively correlated even in recent years, and real wage growth was negative for much of the recent high-inflation episode (figure 1). Average real wage growth was below 1 percent in the year prior to the survey (Van Nostrand, Feiveson, and Sinclair 2023), suggesting that for some sizable share of consumers, purchasing power did decline.

Relatedly, Stantcheva's survey asks the question, "How would you describe the relation between inflation and unemployment?" The answer choices are: when inflation is higher, unemployment is also higher; or, when inflation is higher, unemployment is lower. This question needs an "it depends" option. In theory, it depends on the types of shocks hitting the economy. Empirically, the correlation between inflation and unemployment is weak (figure 2).

Figure 2. Inflation and Unemployment, 1990–2023



Source: Bureau of Labor Statistics, series CPIAUCSL and UNRATE, retrieved from FRED, Federal Reserve Bank of St. Louis.

Note: Figure plots annual CPI inflation against annual unemployment from 1990 through 2023.

Next, no matter what consumers believe about the types of shocks hitting the economy, they could reasonably interpret the survey questions as asking them to think about the ceteris paribus effects of inflation. For example, they are asked, "Has your purchasing power (your real buying power) decreased or increased because of inflation?" "If inflation was lower than it is now, would you say that you would be less stressed, equally stressed, or more stressed than you are now?" All else equal, inflation does reduce purchasing power and increase stress. They are not instructed to think through a full set of counterfactuals.

For other questions, the wording would be difficult even for an economist to interpret. One asks, "If inflation doubled, how long until your wage doubles?" If inflation were to double, say from 3 percent to 6 percent, I think it would take many years for my wage *level* to double, so I am not surprised that consumers also expect it to take a long time.

Finally, survey responses are likely highly influenced by priming and experimenter demand effects. Respondents are asked many questions about the costs of inflation, how it affects them, and why they dislike it. By the time they are asked to rank inflation among economic and social priorities, inflation is at the top of their mind and it is obvious that the experimenter wants them to dislike inflation, so it is almost inevitable that many rank inflation as a top priority. If the entire survey had been about health care, or unemployment, or abortion, those might have ranked higher.

CONCLUSIONS Stantcheva notes that "people scarcely acknowledge any positive impacts from inflation." The way I think of it, inflation itself does not inherently have positive impacts. But stabilizing aggregate demand, which sometimes requires allowing temporarily higher inflation, does have positive impacts. Inflation is often a side effect of policies that people do like, such as fiscal stimulus in a pandemic. It is perfectly reasonable for people to report that they dislike the side effect, even if they would dislike the counterfactual (no stimulus and low inflation) even more. It is also reasonable for people to strongly dislike, and for the media to fixate on, inflation that results from actual or perceived policy errors or political incompetence.

Understanding preferences and beliefs about inflation is certainly an important part of understanding the costs of inflation, and Stantcheva's new data set will be a valuable tool for researchers in this area. Stantcheva's paper complements related work; for example, using data from the World Values Survey from forty-two countries, Magud and Pienknagura (2024) show that consumers around the world express more concern for price stability if they have lived through high-inflation episodes. Other complementary work is by Afrouzi and others (2024), who survey US consumers about their longer-run inflation preferences.

REFERENCES FOR THE BINDER COMMENT

Afrouzi, Hassan, Alexander Dietrich, Kristian Myrseth, Romanos Priftis, and Raphael Schoenle. 2024. "Inflation Preferences." Working Paper 32379. Cambridge, Mass.: National Bureau of Economic Research.

Bailey, Martin J. 1956. "The Welfare Cost of Inflationary Finance." *Journal of Political Economy* 64, no. 2: 93–110.

Binder, Carola. 2024. *Shock Values: Prices and Inflation in American Democracy*. Chicago: University of Chicago Press.

Blanchard, Olivier, and Stanley Fischer. 1989. *Lectures on Macroeconomics*. Cambridge, Mass.: MIT Press.

Eggertsson, Gauti B., and Donald Kohn. 2023. "The Inflation Surge of the 2020s: The Role of Monetary Policy." Hutchins Center Working Paper 87. Washington: The Brookings Institution.

Federal Reserve Board of Governors. 2012. "Statement on Longer-Run Goals and Monetary Policy Strategy." Washington: Author. https://www.federalreserve.gov/monetarypolicy/files/FOMC_LongerRunGoals_201201.pdf.

Federal Reserve Board of Governors. 2015. "FOMC Meeting Statement." Press release, December 16. https://www.federalreserve.gov/monetarypolicy/fomcpresconf20151216.htm.

Federal Reserve Board of Governors. 2020a. *Fed Listens: Perspectives from the Public*. Washington: Author. https://www.federalreserve.gov/publications/files/fedlistens-report-20200612.pdf.

Federal Reserve Board of Governors. 2020b. "2020 Statement on Longer-Run Goals and Monetary Policy Strategy." https://www.federalreserve.gov/monetarypolicy/review-of-monetary-policy-strategy-tools-and-communications-statement-on-longer-run-goals-monetary-policy-strategy.htm.

Magud, Nicolas E., and Samuel Pienknagura. 2024. "Inflated Concerns: Exposure to Past Inflationary Episodes and Preferences for Price Stability." Working Paper 2024/040. Washington: International Monetary Fund.

Mankiw, N. Gregory. 1997. "Comment on 'Why Do People Dislike Inflation?' by Robert J. Shiller." In *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer. Chicago: University of Chicago Press.

Romer, Christina D., and David H. Romer. 1997. "Introduction." In *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer. Chicago: University of Chicago Press.

Samuelson, Paul A. 1938. "A Note on the Pure Theory of Consumer's Behaviour." *Economica* 5, no. 17: 61–71.

Shiller, Robert J. 1997. "Why Do People Dislike Inflation?" In *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer. Chicago: University of Chicago Press.

Shiller, Robert J. 2011. "The Neuroeconomics Revolution." Project Syndicate, November 21. https://www.project-syndicate.org/commentary/the-neuroeconomics-revolution.

Van Nostrand, Eric, Laura Feiveson, and Tara Sinclair. 2023. "The Purchasing Power of American Households." US Department of the Treasury, December 14. https://home.treasury.gov/news/featured-stories/the-purchasing-power-of-american-households.

Varian, Hal R. 2006. "Revealed Preference." In *Samuelsonian Economics and the Twenty-First Century*, edited by Michael Szenberg, Lall Ramrattan, and Aron A. Gottesman. Oxford: Oxford University Press.

Wen, Yi. 2010. "Why Do People Dislike Inflation?" *Economic Synopses* 15. https://files.stlouisfed.org/files/htdocs/publications/es/10/ES1015.pdf.

COMMENT BY

YURIY GORODNICHENKO In a seminal contribution, Shiller (1997) used a series of surveys to understand why people strongly dislike inflation while economists have relatively benign views on inflation. That paper presented a puzzle, but as Mankiw (1997) observed, it was not clear what one should do about this puzzle. Would people be more relaxed about inflation if inflation stayed low and stable for a long time? Would the results in Shiller (1997) carry to other environments? What do people think about inflation now, after a recent short-lived but significant spike in inflation? Stantcheva presents a highly timely study that sheds more light on these important questions.

She finds that, consistent with Shiller (1997), people intensely dislike inflation and rank inflation as one of the most pressing issues in the country. Several key features stand out. First, people interpret inflation as a bad state of the world. For example, they think that inflation is positively correlated with unemployment (i.e., inflation is stagflationary). In contrast, economists (professional forecasters) generally see a negative correlation between inflation and unemployment, which is consistent with a Phillips curve and business cycles driven by demand-side shocks. Second, people take a partial equilibrium approach to inflation: they believe that inflation reduces their purchasing power. Furthermore, few households name monetary or fiscal policy as the source of inflation. Instead, the common answers include energy and food costs, which are often only proximate causes of price increases. On the other hand, economists generally believe that moderate levels of inflation do not affect real wages and that expansionary monetary policy and fiscal imbalances are the key sources of inflation (e.g., Milton Friedman observed, "Inflation is always and everywhere a monetary phenomenon" [1994, 49]). Third, people see no benefits of positive inflation and, if anything, think about inflation as a zero-sum game where inflation redistributes resources from one group of economic agents to another. Again, this contrasts with economists' conviction that inflation can be beneficial (e.g., reduce unemployment and avoid deflationary spirals). Furthermore, none of the costs of inflation (e.g., price dispersion, menu costs) that are emphasized by economists are systematically mentioned by people. Fourth, people often "personalize" blame for inflation (i.e., a specific person is responsible for inflation) while economists take a more nuanced view. Finally, people's take on inflation is strongly colored by their political leanings. Republicans blame incumbent Democrats for inflation in recent years, and one may expect the Democrats would blame Republicans if Republicans were in power. Political polarization thus translates to extreme views about economic issues as well.

One can conclude that—to paraphrase Mankiw (1997)—economists are not people and people are not economists. The differences are so stark that one may be tempted to assert that: (1) people do not know what they are talking about; (2) their views on inflation do not affect their choices; and

(3) rational agents such as financial markets and managers of firms are the relevant group. The economics profession adopted various combinations of these reactions and thus largely ignored what people think about inflation. This strikes me as a wrong response. First, Stantcheva's paper and other surveys document that although inflation is a confusing subject for many households, many people in a low-inflation environment (where incentives to understand inflation are weak) provide imperfect but *close enough* definitions of inflation. For example, Stantcheva finds that about 50 percent of respondents in her survey of US households give a reasonable definition of inflation. Other studies document that this fraction is higher for more financially literate households and for households who have experienced significant inflation in the past.[1] These results suggest that people have at least some idea about what inflation means.

Second, the mapping from what people think about inflation and how they act on their views can be indeed complex and establishing causal links is difficult. However, recent studies combining randomized controlled trials (RCTs), surveys, and administrative data document that exogenous variation in inflation expectations of households and firms affects their choices. For example, Coibion, Gorodnichenko, and Weber (2022) provide randomly chosen households with publicly available information about inflation (e.g., the Federal Reserve's inflation target) to create exogenous variation in their expectations and then use this exogenous variation to show that raising inflation expectations *lowers* spending on durable goods (which is consistent with households' stagflationary view on inflation). In a similar spirit, Coibion, Gorodnichenko, and Ropele (2020) document that exogenously higher inflation expectations cause firms to raise their prices. Hence, it is true that survey measures of inflation expectations of households and firms have responses looking strange to economists, but these survey responses do contain useful information and economic agents act on their beliefs.

Third, financial markets are clearly much more informed than households, but the distance between firm managers and households is not as large as one may think. Casual observations of what captains of the industry opine on inflation suggest that inflation can be a confusing subject for them too.[2] More systematic analysis of firms' inflation expectations (e.g., Candia, Coibion, and Gorodnichenko 2024) suggests that firms' expectations fall somewhere between households' and professional forecasters'. For example,

1. See D'Acunto, Malmendier, and Weber (2023) for a survey.

2. For example, on October 22, 2022, Elon Musk declared in an interview, "There's more deflation than inflation" (Henney 2022, par. 3). According to the US Bureau of Labor Statistics, the Consumer Price Index inflation rate in October 2022 was 7.8 percent.

figure 1 shows that although managers have less disagreement than households and more than professional forecasters, firms' expectations appear to be as unanchored as households' during the 2021–2023 inflation spike. Similar to households, managers appear to rely on gas prices and personal shopping experience when they form their inflation expectations (e.g., Kumar and others 2015). Thus, one may expect that Stantcheva's findings for households should largely apply to firms too.

Mankiw (1997, 68) asked a key question, "If ignorance [about inflation] is in fact pervasive, how should that fact alter economic theory and policymaking?" He suggested that the response may range from "do nothing" (Sherlock Holmes did not know that the earth revolved around the sun because it was not important for his daily life) to "take it seriously" ("inflation is undesirable precisely because it is misunderstood" [ibid.]). Stantcheva's survey results and other evidence make me think that one should take it seriously. To support this view, let me provide three reasons.

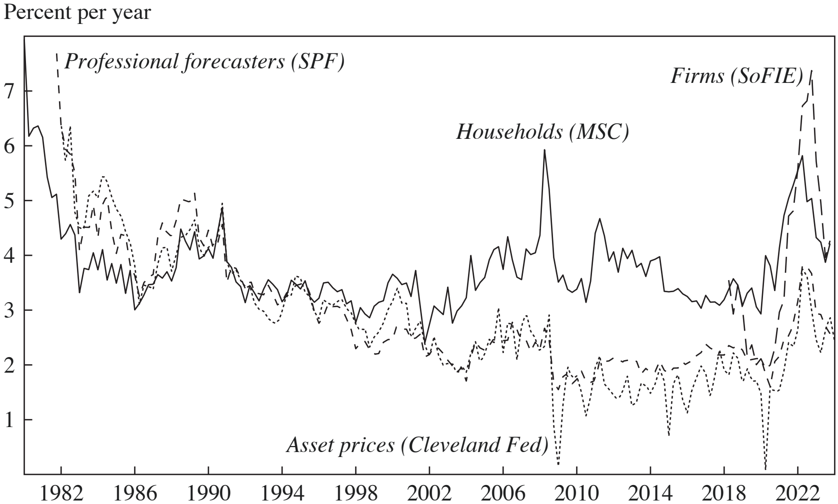
First, New Keynesian macroeconomics shows that the central bank should minimize variance of output gap X_t and inflation π_t with some weight ω on the latter, that is, $\text{var}(X_t) + \omega \times \text{var}(\pi_t)$. Theory often implies that the weight on inflation should be very high (100 or above). This very high weight makes many economists uncomfortable, and it is not unusual to see that much lower ad hoc weights such as $\omega = 1$ are used in applied work. In other words, economists have a hard time making inflation a priority. People, on the other hand, appear to want low inflation as a high priority for central banks (that is, $\omega \gg 1$). Consistent with Stantcheva's evidence, Afrouzi and others (2024) find that households' preferred inflation target is zero. Although one can make a strong theoretical argument for why zero inflation may be a poor choice, it could be politically imprudent to ignore public opinion on this matter and raise the inflation target from 2 to say 4 percent or more.

Second, central banks employed a variety of strategies to raise inflation (and inflation expectations) after the global financial crisis in 2007–2009 to stimulate aggregate demand. For example, Mario Draghi (2015) explained, "When inflation expectations go up with zero nominal rates, real rates go down. When real rates go down, investments and the economic activity improves. That's the reasoning [of QE]." However, if households view inflation as stagflationary, raising inflation can make households reduce consumer spending rather than increase. In other words, strategies focused on raising inflation expectations can backfire.

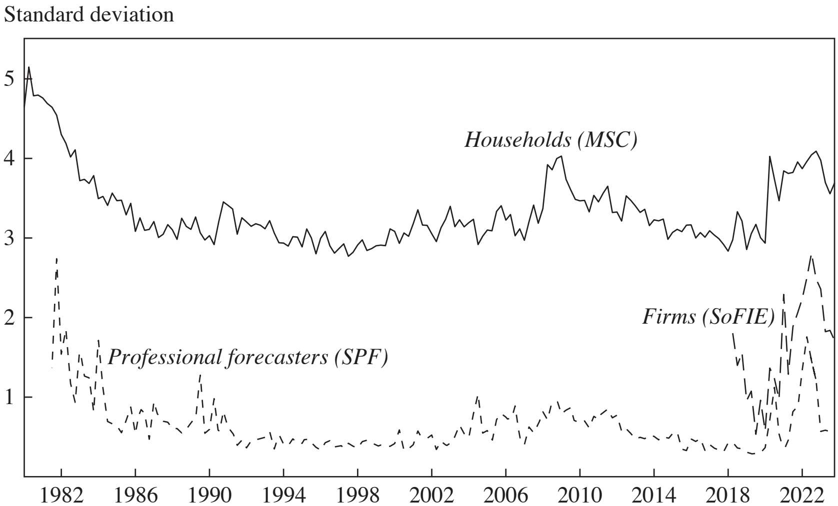
Third, to be effective, certain policy tools require economic agents to understand general equilibrium effects and to have the ability to iteratively eliminate dominated strategies. For example, price-level targeting requires

Figure 1. One-Year-Ahead Inflation Expectations for Different Agents

Panel A: Mean forecast



Panel B: Disagreement (standard deviation)



Source: Reproduced from Candia, Coibion, and Gorodnichenko (2024) with permission, copyright Elsevier.

Notes: Financial markets' expectations are from the Federal Reserve Bank of Cleveland, households' expectations are from the Michigan Survey of Consumers (MSC), and professional forecasters' expectations are from the Survey of Professional Forecasters (SPF) run by the Federal Reserve Bank of Philadelphia. Responses of households that are greater than 15 percent or less than –2 percent are excluded. Firms' expectations are from the new survey of CEOs in Candia, Coibion, and Gorodnichenko (2024)—Survey of Firms' Inflation Expectations (SoFIE). Responses that are greater than 15 percentage points or less than –2 percentage points are excluded. All moments are computed using survey weights.

economic agents to understand that above-average inflation today is followed by below-average inflation tomorrow and thus economic agents should not raise prices today (if their prices are sticky). But if economic agents do not have a strong incentive to raise prices today, then the initial inflationary shock has a smaller effect on inflation and thus incentives to raise prices today are even weaker. As a result, price-level targeting can be a highly powerful tool for macroeconomic stabilization. On the other hand, Stantcheva's results suggest that people have a rather partial equilibrium thinking, and we know from other work (e.g., Camerer 1997) that people tend to have relatively low level-*k* thinking. Thus, one may anticipate that price-level targeting can be less effective (and potentially even destabilizing) in practice.

What are the next steps? Is this the beginning of the end for conventional macroeconomics? In my view, Stantcheva's paper marks the end of the beginning for the literature documenting what people think about inflation. Clearly, people do not like inflation, and this can be important for policy and theory. Future work should focus more on understanding what makes inflation so undesirable for people (e.g., general confusion about inflation, inability to hedge against inflation, level versus uncertainty about inflation) and quantifying forces behind this dislike (e.g., one can use hypothetical questions to get quantitative responses). Stantcheva also cuts out work for macroeconomic theorists. For example, what should macroeconomic stabilization policy look like when people have views that are rather different from those of economists? What policy regime (gold standard, inflation targeting, price-level targeting, flexible average inflation targeting, etc.) is better when economic agents have beliefs that we observe in the data? In short, Stantcheva's important study should keep us busy for quite some time, and I look forward to seeing more work in this arena.

REFERENCES FOR THE GORODNICHENKO COMMENT

Afrouzi, Hassan, Alexander Dietrich, Kristian Myrseth, Romanos Priftis, and Raphael Schoenle. 2024. "Inflation Preferences." Working Paper 32379. Cambridge, Mass.: National Bureau of Economic Research.

Camerer, Colin F. 1997. "Progress in Behavioral Game Theory." *Journal of Economic Perspectives* 11, no. 4: 167–88.

Candia, Bernardo, Olivier Coibion, and Yuriy Gorodnichenko. 2024. "The Inflation Expectations of U.S. Firms: Evidence from a New Survey." *Journal of Monetary Economics* 145 (Supplement): 103569.

Coibion, Olivier, Yuriy Gorodnichenko, and Tiziano Ropele. 2020. "Inflation Expectations and Firm Decisions: New Causal Evidence." *Quarterly Journal of Economics* 135, no. 1: 165–219.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2022. "Monetary Policy Communications and Their Effects on Household Inflation Expectations." *Journal of Political Economy* 130, no. 6: 1537–84.

D'Acunto, Francesco, Ulrike Malmendier, and Michael Weber. 2023. "What Do the Data Tell Us about Inflation Expectations?" In *Handbook of Economic Expectations*, edited by Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw. London: Elsevier.

Draghi, Mario. 2015. "Hearing at the European Parliament's Economic and Monetary Affairs Committee (Introductory Remarks and Q&A)." Speech given at the European Parliament, March 23. https://www.ecb.europa.eu/press/key/date/2015/html/sp150323_1.en.html#qa.

Friedman, Milton. 1994. *Money Mischief: Episodes in Monetary History*. New York: Harcourt Brace & Company.

Henney, Megan. 2022. "Elon Musk Accuses Fed of 'Looking in the Rearview Mirror' with Rate Hikes: Tesla CEO Says US Economy in a Period of 'Deflation'." Fox Business, October 20. https://www.foxbusiness.com/economy/elon-musk-accuses-fed-looking-rearview-mirror-rate-hikes.

Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko. 2015. "Inflation Targeting Does Not Anchor Inflation Expectations: Evidence from Firms in New Zealand." *Brookings Papers on Economic Activity*, Fall: 151–208.

Mankiw, N. Gregory. 1997. "Comment on 'Why Do People Dislike Inflation?' by Robert J. Shiller." In *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer. Chicago: University of Chicago Press.

Shiller, Robert J. 1997. "Why Do People Dislike Inflation?" In *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer. Chicago: University of Chicago Press.

GENERAL DISCUSSION Commenting on the finding that low-income individuals have changed their behavior more in response to inflation, Katharine Abraham noted that this does not necessarily imply that when facing the same price increase, low-income individuals are more responsive than high-income individuals. She referred to research by Xavier Jaravel, which suggests that prices for low-income individuals tend to rise more rapidly than prices for high-income individuals.[1]

John Haltiwanger brought up how substitution bias, product turnover, and quality change contribute to the difficulty of accurately measuring inflation—even average inflation. To Abraham's point, this makes the attempt to measure inflation even more cumbersome as the perception of inflation differs across different groups in the population. He pointed out that we don't

1. Xavier Jaravel, "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector," *Quarterly Journal of Economics* 134, no. 2 (2019): 715–83.

have any real-time measurement of inflation and one reason there is heterogeneity in the responses is that just as economists are struggling to measure inflation, so is everyone else.

Elaine Buckberg built upon Abraham's comment and added that higher-income households are more likely to own their own homes and therefore more likely to experience a positive wealth effect in the recent inflation episode due to rent inflation. Buckberg also responded to Yuriy Gorodnichenko's discussion on how the average respondent does not understand that real wages will catch up over time, contending that what consumers are really saying is that it is too painful to wait while wages catch up.

Stan Veuger commented that with enough heterogeneity in inflation across people and goods, we might get to the point where people get more information out of a trip to the store or a conversation with a friend than from federal statistical agencies.

Steven Davis, using Federal Reserve Bank of Atlanta's Wage Growth Tracker and deflating by the Consumer Price Index for All Urban Consumers (CPI-U), stated that the median value of real wages fell 3.3 percent from 2020:Q3 to 2022:Q4 and were still down by 1.2 percent in 2023:Q3.[2] Davis added that if there is inflation inequality, as Abraham pointed out, these calculations understate the extent of real wage declines for some households. Davis commented that because households had recently experienced sizable decline in real wages at the time of the survey, the negative view on inflation expressed by survey respondents is unsurprising. He remarked that although there may be economic benefits to inflation, experiencing the effect on one's purchasing power is still unpleasant. Davis postulated that this recent episode of inflation would influence policy for some time, because the average person will be more averse to inflation for many years ahead.

Greg Mankiw remarked that the inflation referred to in textbooks is purely monetary and a tool for measurement, but the recent episode of inflation could be the result of adverse supply shocks, which do lower real wages. Mankiw agreed with Davis that people tend to refer to their own recent experience with inflation rather than the textbook definition.

Robert Gordon elaborated on Davis's comment and added that, based on his own calculations, productivity growth for the total economy was about 1.0 percent between early 2020 and mid-2023, which means that the difference between real outcomes and what people would expect in the long run was closer to 3 percent. Gordon also pointed out that the inflation episode

2. Federal Reserve Bank of Atlanta, "Wage Growth Tracker," https://www.atlantafed.org/chcs/wage-growth-tracker.

in the past three years and the one in the 1970s and 1980s were both generated primarily by supply shocks.

Christina Romer explained that people can't see the trade-offs between inflation and unemployment because once people are experiencing inflation, those benefits are in the past. Romer also noted that the survey responses pointing to the Biden administration and policies as primary issues acknowledge the link between policy and inflation, but people might not recognize those same policies also reduced unemployment.

Laura Alfaro pointed out that the findings in the paper are supported by evidence from Latin American countries, which were among the first to raise interest rates to fight the recent inflation episode. She added that Latin American countries know from experience that lower-income individuals are disproportionately hurt by high inflation and often blame their government. She noted the discrepancy between the economic theory of the inflation-unemployment trade-off and the experience of people—for most people, there is no sense of a trade-off, rather, they are just able to afford less than they could before.

Veuger warned about the support that he sensed for a zero-inflation policy. He jokingly highlighted that one of the reasons we have independent central banks is to keep inflation well above zero and that inflation would be suboptimally low if elected officials were in charge of setting inflation.

Andrew Atkeson shared that he teaches inflation using a 1933 Pete Smith newsreel to explain President Roosevelt's policy of going off the gold standard, the subsequent inflation, and the benefits from inflation. Atkeson brought up two related questions on whether there is historical evidence that the public reaction to inflation after going off the gold standard was favorable or unfavorable, and whether economists should consider using storytellers to effectively explain the benefits of inflationary policy.

Gordon explained that during the Roosevelt administration, people were enthusiastic about raising inflation because from 1929 to 1942 the correlation between the price level and real GDP was very high. Since this correlation no longer exists, it is not surprising that people today have very different attitudes toward inflation.

Barry Eichengreen commented that the first Gallup poll was conducted in 1935.[3] In response to the question, "What do you think the biggest problem facing the country is?" the top responses were unemployment, the federal

3. Frank Newport, "75 Years Ago, the First Gallup Poll," blog, Gallup, October 20, 2010, https://news.gallup.com/opinion/polling-matters/169682/years-ago-first-gallup-poll.aspx.

budget, and taxes. Inflation did not appear in the top twenty responses to that question.[4] Eichengreen hypothesized that either people had been traumatized by very high unemployment and low inflation during the previous years, or the propaganda used by Roosevelt worked.

Peter Henry elaborated on Atkeson's point and added that Jamaica was able to reduce its inflation rate with a sustained high interest rate policy by implementing a communication policy to educate the population. He also remarked that because only about a third of the US adult population has gone to college, and a much smaller proportion have studied economics, it should come as no surprise that the public in general are not aware of the connection between inflation and unemployment.

Stefanie Stantcheva responded to the comments about providing information and narratives to the public. She argued that even though there are trade-offs, self-interested people will still care about inflation during high-inflation episodes and unemployment when unemployment is high, because the experienced loss is so acute, suggesting a limited role for pedagogical explanations to educate the public.

Buckberg echoed this concern but suggested survey respondents may think that they would prefer low inflation and high unemployment to high inflation and low unemployment if they believe they would not be the ones experiencing unemployment in a high unemployment situation. Buckberg added that the recent experience of inflation taught her that unemployment affects the unemployed and their immediate families, but inflation affects everyone.

Stantcheva agreed with Buckberg and elaborated that inflation is similar to trade in that there are diffused gains but very concentrated losses. When inflation is high, it becomes very salient; and when unemployment is high, unemployment becomes more salient as the high costs of unemployment start to diffuse across the economy. She added that this saliency changes over time, referring to some of her own new work on this topic.

Bruce Fallick said that one reason people might dislike inflation is due to the cognitive load it causes. He noted that high inflation makes it hard for people to judge prices when they are shopping, and he asked if the idea of cognitive load showed up in the survey responses.

4. Gregor Aisch and Alicia Parlapiano, "'What Do You Think Is the Most Important Problem Facing This Country Today?'" *New York Times,* February 27, 2017, https://www.nytimes.com/interactive/2017/02/27/us/politics/most-important-problem-gallup-polling-question.html.

Henry Aaron brought up Daniel Kahneman's findings on loss aversion, noting that if the variance of price changes goes up with the rate of inflation, economists could expect that the population would be less happy than they were beforehand. He added that this, along with the lag in wage increases, causes people to be hit with multiple losses early on. These may be offset as wages catch up but perhaps only partially.

Alan Blinder responded to Gorodnichenko's presentation and his point about the stagflationary view. In people's mind, when it rains, it pours. He mentioned some of his own recent work on the central bank's communication with the public.[5] He stated that one finding in his paper is that the public mostly misunderstand the sign on interest rates, thinking that higher interest rates are inflationary.

Tara Sinclair mentioned a blog post with Eric Van Nostrand and Laura Feiveson, which received some pushback from people on the view that there have been gains in purchasing power.[6] She brought up the idea that people might be imagining a ceteris paribus situation where inflation is lower, but their wages stay the same. Sinclair raised the question of how survey respondents are thinking about the wage process, wage gains, and how much of those gains come from performance rather than a cost-of-living increase. In response, Gordon commented that people look at inflation as taking something away, but they see wage increases as a reward for their own effort, noting that most people do not consider *real* wages.

Robert Hall remarked that the data from this survey could contribute to the current research on the dynamics of the individual households such as consumption patterns.

Maurice Obstfeld conjectured that one's nominal liabilities plausibly affect attitudes toward inflation—a high liability would make inflation seem more desirable. He also brought up an important historical example of high demand for inflation—during the silver agitation in the United States in the nineteenth century, farmers saw inflation as a way to raise agricultural prices and reduce their real debts.

5. Alan S. Blinder, Michael Ehrmann, Jakob de Haan, and David-Jan Jansen, "Central Bank Communication with the General Public: Promise or False Hope?" *Journal of Economic Literature* 62, no. 2 (2024): 425–57.

6. Eric Van Nostrand, Laura Feiveson, and Tara Sinclair, "The Purchasing Power of American Households," US Department of Treasury, December 14, 2023, https://home.treasury.gov/news/featured-stories/the-purchasing-power-of-american-households; and "An Update to 'The Purchasing Power of American Households'," US Department of Treasury, January 25, 2024, https://home.treasury.gov/news/featured-stories/an-update-to-the-purchasing-power-of-american-households.

Stantcheva responded that the survey suggests people do not associate higher inflation with easier debt repayments; rather, respondents indicated that they believed they were going to be poorer and, as a result, meeting debt obligations would be harder—despite the fact that inflation would induce a decrease in the real value of their debts.

Blinder asked if the survey results could help shed light on the public's failure to differentiate between the price level and the rate of change in the price level. He pointed out that a lot of the public's complaints boil down to items costing more now than they did four years ago, and little attention is paid to the fact that the CPI inflation has fallen from 9 percent to about 3 percent.[7]

Jonathan Pingle referred to work done by Steinsson and Nakamura, which distinguishes between periods of inflation characterized by many small increases in prices versus those characterized by larger increases in prices.[8] He postulated that this distinction could help explain the experiences that Robert Shiller encountered relative to Stantcheva's findings.

Wendy Edelberg posited that while people may not be able to tell the difference between 2 percent and 3 percent inflation, the survey does indicate what people's response is when inflation is notably higher. Further, she questioned if there would be different policy outcomes for dealing with inflation if the population was more educated on the topic. She pointed out that she would like to know to what extent the issue at hand relates to the political economy and to what extent it would simply yield different outcomes in economic modeling if people had a more nuanced view on inflation.

Stantcheva responded that economists have a lot to learn from the public's understanding of these issues, and beyond misperceptions among the public, people may be facing constraints that economists are unaware of. She suggested economists keep this in mind.

7. Bureau of Labor Statistics, "12-Month Percentage Change, Consumer Price Index, Selected Categories," https://www.bls.gov/charts/consumer-price-index/consumer-price-index-by-category-line-chart.htm.

8. Emi Nakamura and Jón Steinsson, "Five Facts about Prices: A Reevaluation of Menu Cost Models," *Quarterly Journal of Economics* 123, no. 4 (2008): 1415–64 (and the supplement, which is available at https://eml.berkeley.edu/~enakamura/papers/fivefactssupplement.pdf); "Monetary Non-neutrality in a Multisector Menu Cost Model," *Quarterly Journal of Economics* 125, no. 3 (2010): 961–1013; and "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications," *Annual Review of Economics* 5 (2013): 133–63.

ANDREW ATKESON
University of California, Los Angeles

STEPHEN KISSLER
University of Colorado, Boulder

The Impact of Vaccines and Behavior on US Cumulative Deaths from COVID-19

ABSTRACT We estimate that the combination of changes in behavior to slow the spread of COVID-19 and the delivery of vaccines to a substantial majority of the American population by mid-2021 saved close to 800,000 American lives relative to what would have occurred had vaccines not been developed. We argue that the duration and magnitude of this behavioral response—and thus its overall success in delaying infections—came as a surprise, relative to both our historical experience with pandemic influenza and to model-based projections based on that experience. Thus, we take from our experience with COVID-19 over the past four years the important public health lesson that behavior change can be a powerful force for slowing the spread of a dangerous infectious respiratory disease for a long time. At the same time, these behavioral changes to slow the spread of COVID-19 came at a tremendous economic, social, and human cost. To avoid similar pain from mitigation in the next pandemic, we argue that we need to make investments now not only in vaccine development, but also in data infrastructure so that we can precisely target behavior-oriented mitigation efforts to minimize their economic and social impacts in the next pandemic.

Conflict of Interest Disclosure: Stephen Kissler has served as a consultant for ModernaTx, advising on infectious disease modeling for vaccine strategy. The authors did not receive financial support from any firm or person for this paper or, other than the aforementioned, 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 2024: 67–112

Starting in March 2020, the American public undertook massive changes in behavior in response to the threat from COVID-19.[1] These behavioral changes arose partly in response to public mandates and partly as a spontaneous private reaction to this new disease threat. These public and private disease mitigation efforts succeeded in slowing the transmission of SARS-CoV-2 to a remarkable extent through 2020 and well into 2021, by which time effective vaccines had been developed and delivered to much of the American population.

As a result of these mitigation efforts, a large majority of Americans were able to get vaccinated for COVID-19 before experiencing their first infection. We document this using nationwide serology data, which lets us estimate the cumulative number of infections and vaccinations over time. Population-level data on vaccine efficacy indicate that this success in delivering vaccines to many Americans prior to their first SARS-CoV-2 infection substantially reduced the infection fatality rate (IFR) these Americans suffered when they did contract COVID-19.

In this paper, we use these observations, together with a structural epidemiological model, to argue that the combined success in slowing SARS-CoV-2 transmission through behavior change and the widespread delivery of vaccines saved close to 800,000 American lives.

We argue that relative to historical experience with pandemic influenza and modeling based on this experience, this public health success was a surprise. As of March 2020, it was not at all clear that it would be possible to slow the spread of SARS-CoV-2 long enough to develop vaccines and deliver them to the American population in time to save lives. We see the success of behavior-based mitigation of SARS-CoV-2 transmission as one of the most important public health lessons of this pandemic—it is, in fact, possible to slow the spread of a dangerous respiratory disease for quite a long time.

But, at the same time, these mitigation efforts came at a tremendous economic, social, and human cost. To avoid similar pain from mitigation in the next pandemic, we argue that we need to make investments now not only in vaccine development, but also in data infrastructure so that we can precisely target mitigation efforts to minimize the economic and social impacts of mitigation with the next pathogen. One might think of these investments in data infrastructure as similar in spirit to the huge investments

1. In what follows, we use the term "COVID-19" to refer to the disease caused by the SARS-CoV-2 virus.

made in the infrastructure to gather economic data after World War II to better guide economic policy. For population-level infectious disease mitigation policies to be effective at low economic and social cost, they need to be guided by detailed real-time epidemiological, demographic, and behavioral data, which are only available in a crisis if one is prepared in advance to gather such data.

I. Our Estimate of Lives Saved

We estimate that behavioral mitigation and vaccination together saved close to 800,000 American lives between February 15, 2020 and February 15, 2024. This estimate is based on three data sources: serology data capturing immunity derived from SARS-CoV-2 infection and COVID-19 vaccination in the American population, data on the dynamics of COVID-19-associated deaths, and linked vaccine and COVID-19 mortality data from thirty US states. We describe the construction and interpretation of these data in section II.

Our estimate rests on two central premises. First, due to the immune evasion capabilities of SARS-CoV-2, the overwhelming majority of Americans would have become infected with SARS-CoV-2 by February 2024 under any realistic vaccination and behavioral mitigation scenario. Second, the health risk of a person's first infection, when unvaccinated, is vastly higher than one's risk after having been vaccinated or previously infected. Thus, the benefit of behavioral mitigation and vaccination came principally from vaccinating individuals before their first SARS-CoV-2 infection. The serology data indicate that slightly more than two-thirds of the US population were vaccinated prior to their first infection with SARS-CoV-2; it is this group that principally contributes to our estimate of lives saved.

In support of these premises, an estimated 94 percent of Americans had been infected with SARS-CoV-2 by late 2022, despite the behavioral mitigation and vaccine uptake in the preceding years (Klaassen and others 2023). Population-level data on COVID-19 mortality for those who had been vaccinated versus those who had not been vaccinated gathered from thirty US states with linked mortality and vaccine data are consistent with the view that COVID-19 was extremely dangerous for those who contracted it for the first time without protection from vaccines. For those contracting COVID-19 after vaccination or prior infection, the disease is much less dangerous.

Table 1. Model-Implied Cumulative COVID-19 Deaths

Baseline and alternative scenarios	
Baseline behavior and vaccines	1,180,000
Baseline behavior, no vaccines	1,979,000
No mitigation with vaccines	3,345,000

Source: Authors' calculations.

Based on these premises, we construct both a back-of-the-envelope calculation of the lives saved by mitigating behavior and vaccines and an estimate from a structural epidemiological model that considers behavior, decline in the COVID-19 IFR over time, and waning immunity against both reinfection and severe disease. The back-of-the-envelope calculation, which conjectures that the 68 percent of Americans that managed to get vaccinated prior to their first SARS-CoV-2 infection would have suffered an IFR four times higher had they not been vaccinated, leads us to an estimate of 845,000 lives saved.

We develop a full structural model to delve a bit deeper into this calculation and set ourselves up for conducting counterfactual exercises. The model combines a fairly detailed epidemiological description of the various variants of SARS-CoV-2 that have appeared over the past four years with a simple model of how mitigating behavior reacts to the rise and fall of daily deaths from the disease as well as parameters governing the administration of vaccines.[2] We argue that this model fits both the dynamics of the data on COVID-19 deaths and the dynamics of the serology data on infections and vaccinations quite well. Our model's implications for cumulative COVID-19 deaths from February 15, 2020 through February 15, 2024 are shown in the first row of table 1.

We simulate the model with vaccines turned off to arrive at a counterfactual prediction for the dynamics of COVID-19 deaths in the absence of vaccines, with results for cumulative mortality in this counterfactual reported in line 2 of table 1. The use of the full structural model with its added detail delivers our preferred estimate of just under 800,000 lives saved as the difference between cumulative deaths reported on line 2 and line 1.

We then draw out four lessons for future pandemics from these data and our counterfactual modeling exercises.

2. We have presented versions of this model in earlier work, including Atkeson (2021a, 2021b, 2023b).

I.A. Lesson 1

First, we argue that it was the combination of mitigating behavior and vaccines together that saved lives.

To illustrate this point, we conduct two counterfactual model simulations. We simulate our model with its baseline specification of mitigating behavior but without vaccines. Without vaccines, behavior alone would have postponed infections, but in the end, nearly everyone would have been infected and subject to a high IFR from that first infection.

We then simulate our model with vaccines distributed starting at the end of December 2020 but with no mitigating behavior before that time. We report our model-implied cumulative death toll for this scenario in the third row of table 1. In this counterfactual simulation, we see that, without a behavioral response, vaccines would have come too late to save lives. Our model implies that cumulative COVID-19 deaths would also have been substantially higher in this scenario without mitigation because our serology and deaths data imply that COVID-19 was substantially more dangerous in 2020 than in 2021, and most infections in this scenario would have occurred in 2020.

One might be tempted to use this scenario of an unmitigated epidemic as a benchmark against which to argue that the combination of vaccines and behavior together saved over 2 million lives. We argue that such a comparison would be an overstatement as it seems highly implausible that there would be no private efforts to avoid transmission even in the absence of any public mitigation policies. The model simulation of an unmitigated epidemic has the daily death toll peaking at over 60,000 deaths per day. It seems highly likely that people would have reacted on their own to such an outcome even in the absence of any public policies toward the epidemic.

I.B. Lesson 2

This success of delaying infections for many months through changes in behavior was a surprise relative to historical experience and modeling of pandemic influenza.

We take as the strongest piece of evidence in favor of this claim the conclusion of Ferguson and others (2006), a prominent study of mitigation options for a pandemic influenza in the United States, regarding the timing of administration of vaccines that these vaccines would have "almost no effect" (451) if started after 120 days after the first worldwide case because at this time horizon, they would be too late to save lives. Clearly, mitigation of COVID-19 bought us many more than 120 days for vaccines to have a significant impact on COVID-19 cumulative mortality.

I.C. Lesson 3

To a remarkable extent, this strong behavioral response to COVID-19 through 2020 and 2021 was universal across all fifty states.

Certainly, there are significant differences in cumulative mortality from COVID-19 across states, but we argue that the outcomes across US states have much more in common than any of them (except New York City) have with the predicted impact of an unmitigated epidemic. We take these common dynamics of COVID-19 across states as strong evidence of the importance of an endogenous behavioral reaction to current disease incidence as predicted by many economic models.[3]

And yet, this observation leads us to our fourth lesson.

I.D. Lesson 4

It is unclear what behavioral reaction to expect in response to the next epidemic.

Epidemiologists have noted the impact of changes in behavior on the dynamics of prior epidemics, particularly in attenuating the initial phase of exponential growth of infections predicted by simple epidemiological models.[4] But figuring out how to predict the quantitative impact of such changes in behavior and how private behavior will respond to public health measures has proved an unsolved challenge.[5]

We see any successful theory of behavior as having to confront a wide range of data across different epidemics. For example, as noted above, the success of public and private changes in behavior in slowing the spread of COVID-19 came as a surprise relative to historical experience. And yet New York City suffered a terrible first wave of deaths from COVID-19 early in the pandemic largely due to a delayed reaction to the disease despite clear warnings from the Italian experience a few weeks earlier. Somehow the evidence of COVID-19 deaths in New York City seemed to have a much bigger impact on behavior elsewhere in the United States than did the European experience despite objective evidence that air travel links were likely to spread the disease across the globe.

3. See, for example Atkeson (2021b), Gans (2022), and Atkeson, Kopecky, and Zha (2024) and the papers cited therein.

4. See, for example, Chowell and others (2016) and Eksin, Paarporn, and Weitz (2019).

5. See, for example, Ferguson (2007) and Funk and others (2015).

Of particular concern is the question of how our collective experience with COVID-19 over the past four years will influence behavioral responses to the next pandemic for perhaps a generation or more.

The remainder of our paper is organized as follows. In section II, we review the data used in our study. In section III, we summarize the main features of our structural epidemiological model. In section IV, we present our main results and the four main lessons we take away from these results. In section V, we lay out more specifically the types of investments in data infrastructure that we believe would be useful in preparing to do more targeted mitigation in the next pandemic. Finally, in section VI, we conclude.

In section A of the online appendix, we compare the implications of our model to other estimates of lives saved in the literature. In section B of the online appendix, we look more closely at the cross-section of outcomes for COVID-19 cases, vaccinations, and deaths across US states. We use our model to argue that the range of outcomes observed are consistent with plausible variation either in the strength of the behavioral reaction or in state-specific structural factors having an impact on transmission rates. Disentangling the importance of these factors as well as state-level variation in IFRs is something we leave for future research. In section C of the online appendix, we give a full description of our model and its parameters.

II. Serology and Mortality Data

In this section, we review the serology data and the data on mortality from COVID-19 that we use in choosing parameters for our model and constructing our estimate of the impact of behavior and vaccines on cumulative mortality from this disease.

II.A. Serology Data

The serology data we use are drawn from two surveys.

As described in Jones and others (2021), the Centers for Disease Control and Prevention (CDC) measured SARS-CoV-2 seroprevalence (the population-level prevalence of immune markers in the blood) from 2020–2022 by testing for antibodies against two distinct viral antigens in samples from blood donors. One of these antibody types (against type S antigen) is generated in response to either a prior infection or vaccination. The other antibody type (against type N antigen) is generated only in response to prior infection. Thus, with some caveats, the pair of positive or negative results for each sample allows one to measure whether the individual making the blood donation had been previously infected (with or without vaccination), vaccinated without prior infection, or neither vaccinated nor previously infected.[6] We refer to this survey as the Blood Donor Survey.[7]

As described in Bajema and others (2021), serology data were also collected from samples from commercial blood testing labs. These data measure only whether the person giving the sample had previously been infected. We refer to this as the Commercial Lab Survey.[8]

We note that these serology surveys were drawn from different convenience samples—one a sample of blood donors and the other a sample of those having blood drawn as part of their medical checkups or care. We check for consistency of the measure of those infected across these two sources. Unfortunately, no serology data from a sample designed to be representative of the population are available.

In figure 1, we show the results of the Blood Donor and Commercial Lab serology surveys at the national level for the overall population. The crosses show estimates from the Blood Donor Survey of the cumulative percentage of the population that had experienced infection by the survey date (showing a response to the N antigen). The dots show estimates from the Commercial Lab Survey of the cumulative percentage of the population that had experienced infection by the survey date. We see that the two serology surveys give consistent estimates for the percentage of the population infected at least through the first Omicron wave in early 2022.

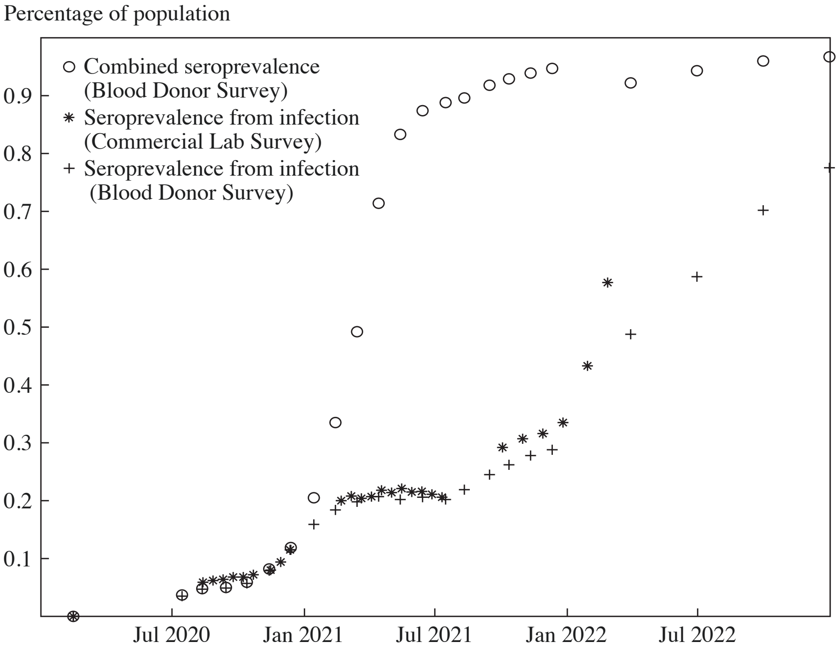
The circles in figure 1 show estimates from the Blood Donor Survey of combined seroprevalence. That is, it adds to the percentage showing a response of the N antigen, the percentage of those showing a response to the S antigen but not the N antigen. This additional group is presumed to be vaccinated but not yet infected; these circles show the sum of those with measurable antibodies from infection (whether or not they also have

6. Such caveats include waning immunity, which can cause a previously infected or vaccinated person to test negative on one or both of the antigen tests, and heterogeneity in the immune response, in which a person can mount an abnormally low immune response to the N antigen despite being infected. See also Ong and others (2021).

7. CDC, "2020–2021 Nationwide Blood Donor Seroprevalence Survey Infection-Induced Seroprevalence Estimates," https://data.cdc.gov/Laboratory-Surveillance/2020-2021-Nationwide-Blood-Donor-Seroprevalence-Su/mtc3-kq6r/about_data; "2022–2023 Nationwide Blood Donor Seroprevalence Survey Combined Infection- and Vaccination-Induced Seroprevalence Estimates," https://data.cdc.gov/Laboratory-Surveillance/2022-2023-Nationwide-Blood-Donor-Seroprevalence-Su/ar8q-3jhn/about_data.

8. CDC, "Nationwide Commercial Laboratory Seroprevalence Survey," https://data.cdc.gov/Laboratory-Surveillance/Nationwide-Commercial-Laboratory-Seroprevalence-Su/d2tw-32xv/about_data.

Figure 1. National-Level Results of the Blood Donor and Commercial Lab Serology Surveys for the Overall Population



Source: CDC.

been vaccinated) and those with antibodies from vaccination but not from infection.

Several features of these serology survey data stand out. First, we see that the estimated cumulative percentage of the population infected as of January 2021 was quite low—well below 20 percent for the overall population. That is, efforts to slow the spread of SARS-CoV-2 through 2020 appear to have succeeded.

We see from the gap between the circles and the crosses/dots that the rapid deployment of vaccines succeeded in vaccinating a large portion of the population prior to first infection by the late summer of 2021. Consistent with this rapid deployment of vaccines in the first half of 2021, we see slow growth in the estimate of cumulative infections between January 2021 and July 2021. From the start of 2021 through the summer of that year, the combination of behavior and vaccinations appeared to be on a path of ending the epidemic.

Unfortunately, due to the combination of new variants (Delta and Omicron) and waning of the immunity provided by vaccines and prior infection, in the fall of 2021, we see infections continue to rise, particularly so in 2022. Given that variants of Omicron have continued to show the ability to infect those who had previously been vaccinated (and reinfect those with prior infections), it is likely that by early 2024, an overwhelming majority of the population has experienced a COVID-19 infection. Considerations of herd immunity that were prominently discussed early in the pandemic have turned out ex post not to be relevant due to a combination of immune evasion by new variants and waning immunity.

II.B. Mortality Data

We now turn to our data on mortality from COVID-19. We draw these data from the CDC's COVID Data Tracker website.[9] This data set counts deaths from COVID-19 at both the national and state levels, with deaths for New York City broken out separately.

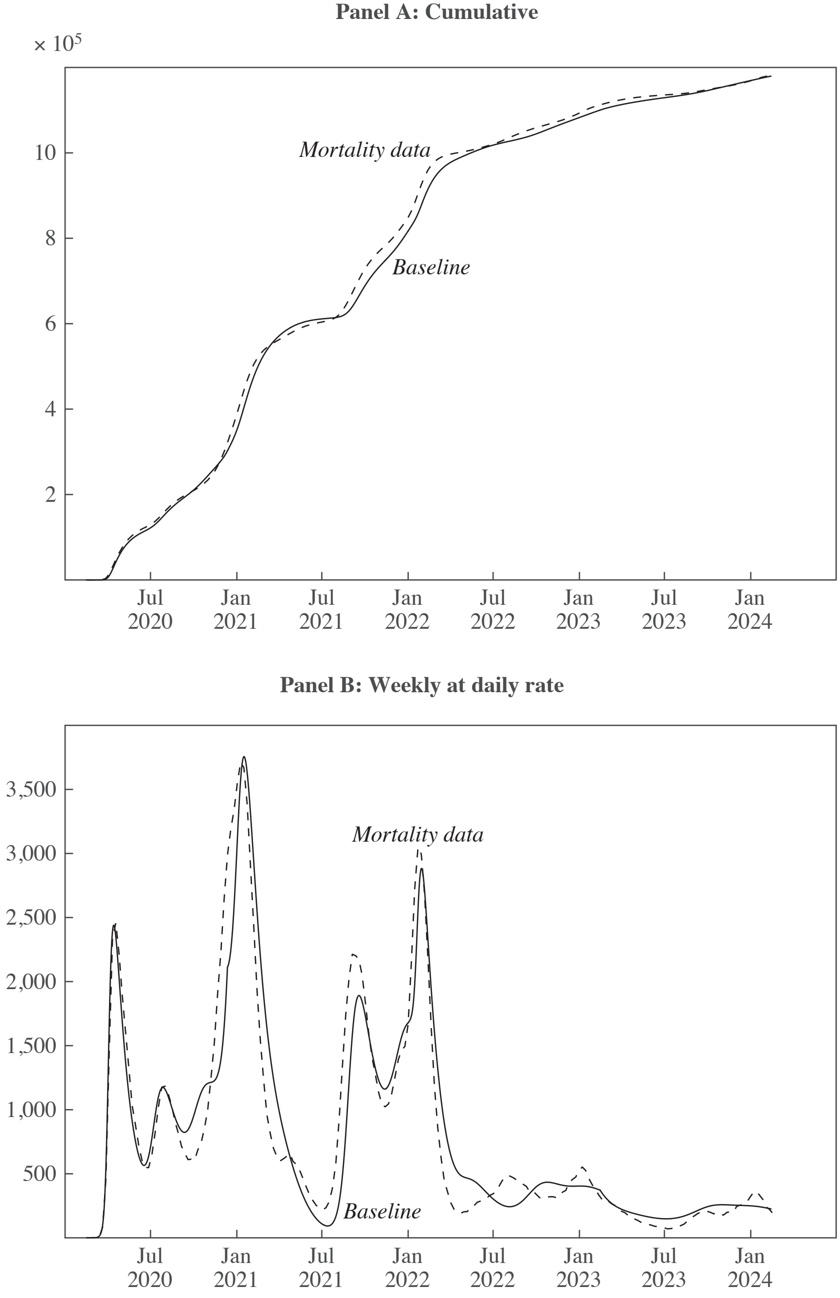
Figure 2 shows cumulative and weekly COVID-19 deaths at a daily rate for the United States from February 2020 to February 2024 in panels A and B, respectively. The mortality data are shown as dotted lines. The outcomes predicted by our baseline model simulation are shown as solid lines.

As shown in this figure, the COVID-19 epidemic in the United States has played out in a series of waves, particularly over the first two years of the epidemic. While these waves garnered considerable attention at the time, what we find most striking about this pattern is that, from very early on in the epidemic, cumulative COVID-19 deaths grew roughly linearly. This linear growth of cumulative deaths is clearly faster in the first two years of the epidemic (from February 2020 through February 2022) than in the second two years of the epidemic.

Why do we find the linear growth of cumulative COVID-19 deaths over the past four years striking? What is missing in figure 2 is any substantial initial period of exponential growth of cumulative deaths as would be predicted by standard epidemiological models for a novel pathogen. To our minds, this observation of linear growth in cumulative deaths sustained over a four-year period is one of the most remarkable features of the COVID-19 epidemic in contrast with historical experience with influenza

9. The data can be downloaded from CDC, "Provisional COVID-19 Death Counts, Rates, and Percent of Total Deaths, by Jurisdiction of Residence," https://data.cdc.gov/NCHS/Provisional-COVID-19-death-counts-rates-and-percen/mpx5-t7tu/about_data.

Figure 2. Baseline Model COVID-19 Deaths for the United States (February 2020 to February 2024)



Source: CDC COVID Data Tracker and authors' calculations.

and the predictions of many epidemiological models. In our model, this outcome is attributed to the strength of the public and private behavioral responses to mitigate transmission of SARS-CoV-2.

II.C. Implied Infection Fatality Rates (IFRs)

Note that these serology and deaths data together imply that the IFR for COVID-19 declined over the course of 2020 and 2021 and then again with Omicron. In particular, the Blood Donor and Commercial Lab Surveys give identical estimates that 11.5 percent of the US population had been infected by December 2020. The cumulative COVID-19 death toll by the end of 2020 was close to 390,000. Given a US population of 332 million, this would imply an overall IFR close to 1 percent.

Looking at the same numbers prior to the first big Omicron wave, in November 2021, the Blood Donor Survey estimated that 27.8 percent of the population had been infected and the Commercial Lab Survey estimated that 31.6 percent of the population had been infected, while the CDC estimates that just over 800,000 Americans had died of COVID-19 by the end of November 2021, implying that close to 61 million Americans were infected with SARS-CoV-2 between January 1 and November 2021 (if we take 30 percent infection-induced seroprevalence as our estimate for November 2021). These numbers imply an IFR closer to 0.66 percent over the course of 2021 prior to Omicron. The equivalent numbers after the first large Omicron wave show a substantial further decline in the implied IFR. We use these estimates as a guide for parameterizing the IFR in our model.

II.D. Mortality by Vaccine Status

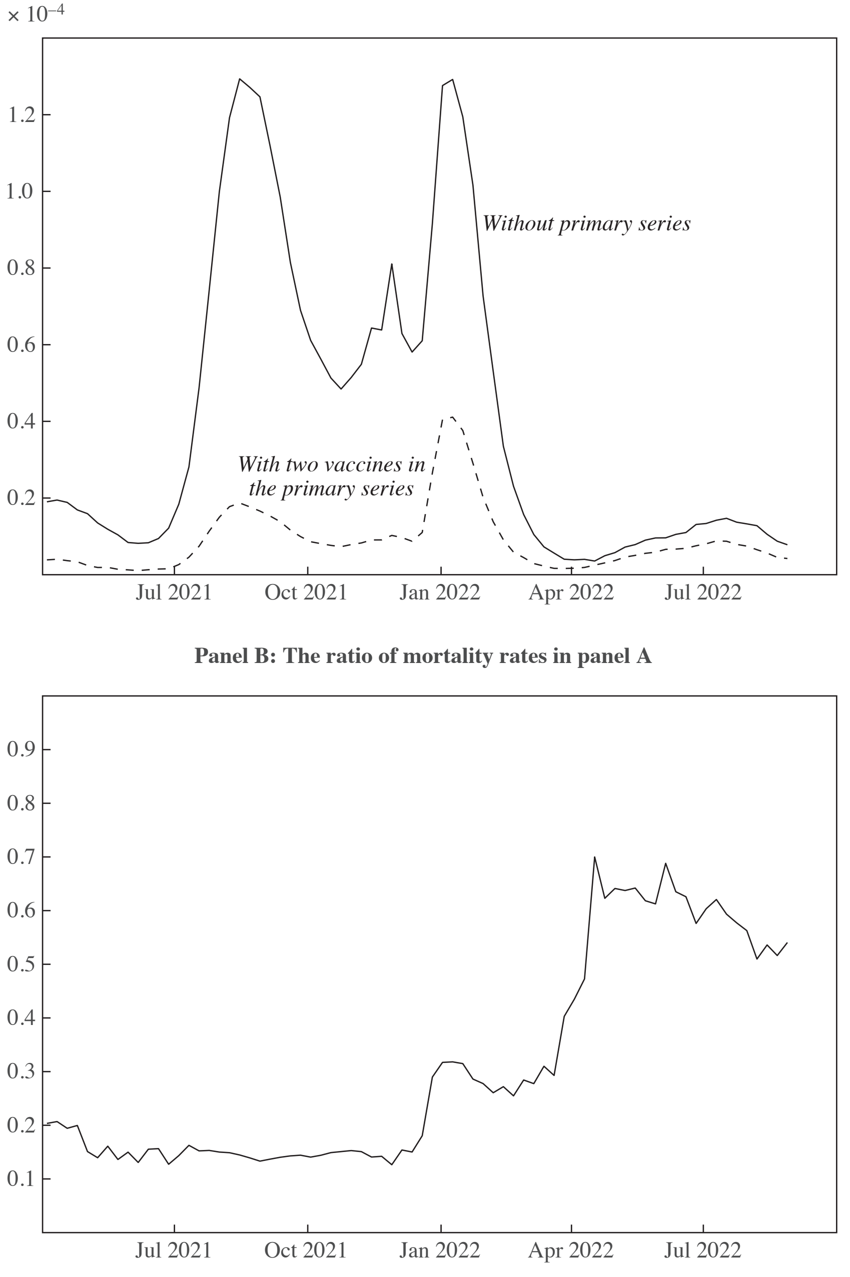
We make use of population-level data on the realized COVID-19 mortality rates of the vaccinated and unvaccinated. As discussed in Jia and others (2023), thirty states of the United States integrated their vaccine databases with their reporting of mortality data. Thus, for these states, on a weekly basis, one can measure the number of COVID-19-related deaths among those who had received the two doses of the primary series of vaccines at least fourteen days before death and COVID-19-related deaths among those who had not received these primary vaccines. The CDC also estimates the number of people in these states in these two groups, and thus one can construct a weekly COVID-19 mortality rate for the vaccinated and unvaccinated populations.

In panel A of figure 3, we show data on the weekly age-adjusted COVID-19 mortality rates for those with two doses of a primary vaccine

Figure 3. Weekly COVID-19 Mortality Rates for the Vaccinated and Unvaccinated Populations

Panel A: Age-adjusted death rates per 100,000 population by vaccination status

Panel B: The ratio of mortality rates in panel A



Source: CDC.

(from the first half of 2021) at least fourteen days prior to death (dashed line) and those without this protection from vaccines (solid line).[10] The dates given on the *x* axis are the year and week number used in the CDC's *Morbidity and Mortality Weekly Report* (*MMWR*). We see in this figure that the weekly mortality rate for the unvaccinated was much higher than for the vaccinated in 2021. After the first big Omicron wave, the weekly COVID-19 mortality rate for the unvaccinated falls to meet the low mortality rate for the vaccinated.

In panel B of figure 3, we show the ratio of these two mortality rates. We see in the panels of this figure that vaccination reduced the COVID-19 mortality rate on the order of 85 percent until the first Omicron wave. After that first Omicron wave, we see that the difference in mortality rates by vaccination status was much smaller. We conjecture, based on the serology data, that this outcome arose as the majority of the unvaccinated had come to have the protection of a prior SARS-CoV-2 infection. Thus, the mortality rates for the unvaccinated fell to a level much closer to that for the vaccinated as both groups were largely protected after this first Omicron wave.

II.E. Waning Immunity and the Long Tail of COVID-19 Deaths

With the emergence of Omicron variants, we have seen that both vaccines and prior infection provide only temporary protection against new infections. As a result, the prevalence of SARS-CoV-2 infections has remained high over the past two years despite the fact that by the end of the first quarter of 2022, the overwhelming majority of the US population had already been vaccinated or experienced a prior SARS-CoV-2 infection or both.

This outcome is a result of two factors. One is that the protection offered by vaccines and prior infection against reinfection wanes over time. The other is that the continual evolution of the virus allows new versions of it to evade immune defenses. After two years of Omicron and three years of experience with mRNA vaccines, it is clear that both processes are at work with COVID-19, but their relative importance is difficult to disentangle.[11]

10. We use the data available at CDC, "Rates of COVID-19 Cases or Deaths by Age Group and Vaccination Status and Booster Dose," https://data.cdc.gov/Public-Health-Surveillance/Rates-of-COVID-19-Cases-or-Deaths-by-Age-Group-and/d6p8-wqjm/about_data.

11. See, for example, Jung and others (2024).

III. Summary Description of the Model

We now present our structural model of the impact of behavior and vaccines on cumulative mortality from COVID-19 in the United States over the period from February 15, 2020 to February 15, 2024. This model extends that in Atkeson (2023b). A full description of this model is given in the online appendix that accompanies this paper.

III.A. Purpose and Fit of the Model

Recall that our estimate of the impact of behavior and vaccines on cumulative COVID-19 mortality is based primarily on an accounting of the number of Americans who were able to get vaccinated prior to their first SARS-CoV-2 infection, and to a lesser degree on an estimate of the benefits of delaying infections due to a decline over time in the IFR of COVID-19. We see this model as a formal accounting device to account for the dynamics of the COVID-19 infection fatality ratio implied by the serology data and the transition of the epidemic toward an endemic steady state.

Thus, while we do not formally estimate the parameters of this model, we do evaluate it as an accounting device on the basis of its fit to the dynamics of SARS-CoV-2 infections and COVID-19 vaccinations at the national level as measured by the serology data in figure 1 as well as the dynamics of deaths from COVID-19 at the national level as shown in figure 2.

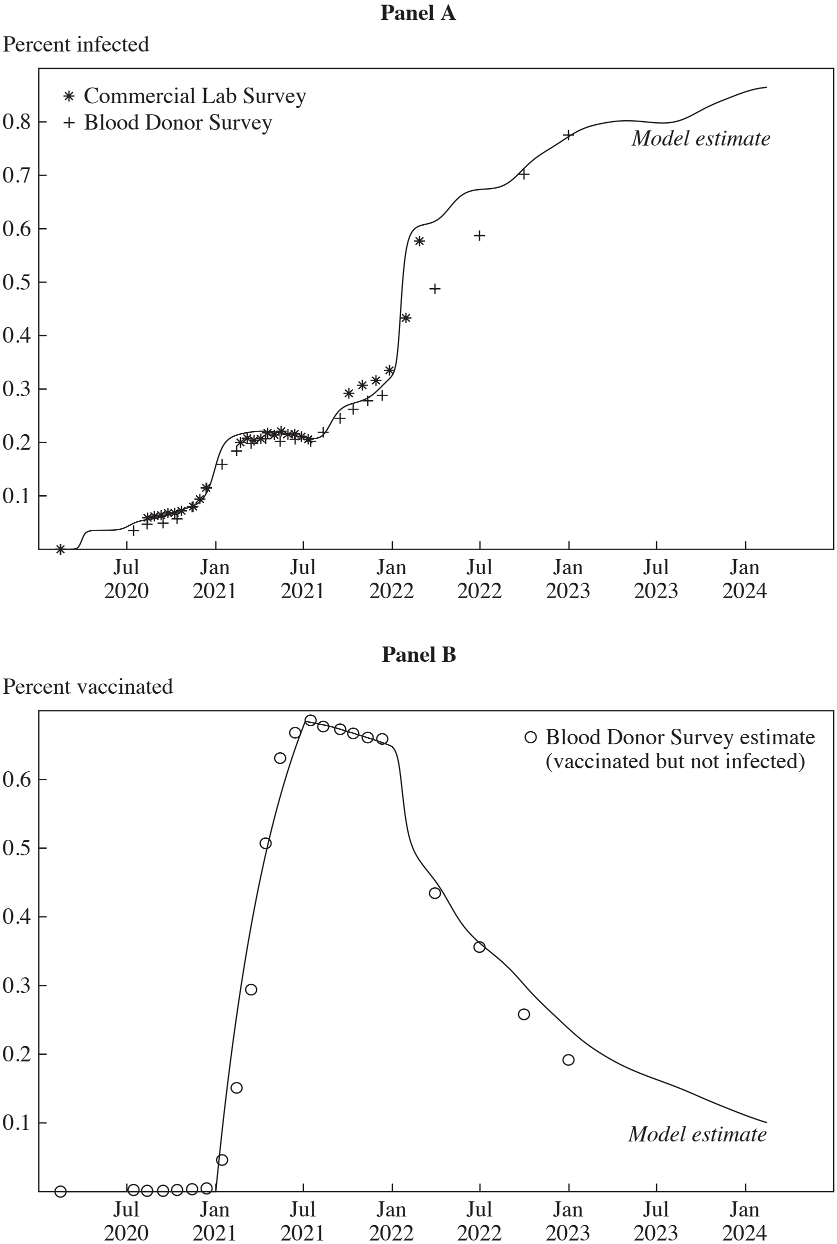
We show the model fit to the serology data in figure 4. Panel A compares the model estimate of the fraction of the population with protection from severe disease due to prior infection, taking into account waning immunity as described below (this fraction can be either vaccinated or not), to the serology data on the fraction of the population showing antibodies from prior infection.

Panel B of figure 4 compares the model estimate of the fraction of the population showing antibodies from vaccination but not prior infection, again taking into account waning immunity as described below, to the serology data on the fraction of the population showing antibodies from vaccination but not prior infection.

In this figure, we see that the fit of the model with its baseline parameters to the dynamics of infections, vaccinations, and deaths is quite good.

We then use this model to assess several counterfactuals to estimate the impact of behavior and vaccines on cumulative mortality from COVID-19 in the United States over the past four years. It is here that the structure of the model is harder to assess as we do not observe these counterfactual outcomes in the data. As we describe the model, we aim to describe what

Figure 4. Fit of the Model to Serology Data



Source: CDC and authors' calculations.

Note: In panel B, the model-implied percentage vaccinated equals to $V(t)/0.75$ where $V(t)$ is the portion of the population with effective protection after vaccination in the model.

features of the data that we do observe allow us to identify the key parameters driving our model's implications for these counterfactuals. In particular, we focus on describing why we have some confidence in our choices for the parameters governing the nature and strength of the behavioral response in the model.

III.B. Model Structure

The model is a susceptible-exposed-infectious-hospitalized-resistant-susceptible (SEIHRS) model with waning immunity and introduction of the Alpha, Delta, and Omicron variants as the epidemic progresses. This model extends the workhorse susceptible-infectious-recovered (SIR) epidemiological model in several dimensions. We explain the reasons for these extensions after reviewing some basic epidemiological concepts.

To begin, a standard SIR model of an epidemic views the population at any point in time as being divided into three categories: susceptible to infection $S(t)$, currently infected and capable of spreading the disease $I(t)$, and resistant to the disease $R(t)$ either from natural immunity (including that induced by prior infection) or from vaccination.

This distribution of characteristics across the population is assumed to evolve over time as follows. Those that are currently infectious, $I(t)$, are assumed to stop being infectious at rate γ per unit time. A fraction η of those who stop being infectious do so because they die. We thus refer to η as the IFR.

Those currently infectious encounter other agents in the population at random and transmit their disease to those agents met at a rate $\beta(t)$ per unit time. We refer to $\beta(t)$ as the *transmission rate*. We allow the transmission rate to depend on factors inherent to the pathogen and the environmental location as indicated by a parameter $\bar{\beta}$ as well as time-dependent factors such as seasonality and behavioral responses.

Since the expected length of time that an infectious agent is expected to be in this state is $1/\gamma$, the average number of agents that an infectious person will transmit their disease to is given by $\beta(t)/\gamma$. Since only fraction $S(t)$ of those agents are actually susceptible to the disease, the expected number of new infections caused by a single infectious agent is given by what is called the *effective reproduction number*:

$$\mathcal{R}_{eff}(t) = \frac{\beta(t)}{\gamma} S(t).$$

Note that the average length of time that an infectious agent remains infectious (here $1/\gamma$) in this model also corresponds to the average length of time between one individual becoming infectious and subsequent infections caused by that individual. This length of time is referred to as the *generation interval*.

The effective reproduction number is related to the SIR model implied growth rate of the fraction of the population that is infectious by

$$\frac{\dot{I}(t)}{I(t)} = \left(\mathcal{R}_{eff}(t) - 1\right)\gamma$$

where $\dot{I}(t)$ denotes the derivative of infections with respect to time.

We note two points from this formula. First, we see that the question of whether the epidemic—in terms of $I(t)$—is growing or shrinking over time is determined by whether the effective reproduction number is above or below one.

Second, the speed of growth of infections per unit time is determined both by the effective reproduction number and the generation interval. Thus, to match data on the growth rate of infections (or deaths) per unit time, one must take a stand on these two parameters. In our model, we hold the generation interval fixed across variants and aim to match the dynamics of weekly deaths in the data with differences in inherent transmissibility of different variants, a seasonal influence on transmissibility, and a behavioral response to the current level of deaths.

We now explain the dimensions in which we extend this simple model and why we do so. We then review our choices for parameter values, with a focus on the generation interval, IFRs, transmission rates, and the impact of behavior on these transmission rates.

We add compartments to the simple SIR model as follows. We add both an exposed state E and the hospitalized state H. Agents in the exposed state have contracted the disease but are not yet infectious. This is a common modification of the SIR framework. Inclusion of this state enriches the dynamics of initial growth of the epidemic. We describe below the purpose of the hospitalized state H. We also add a vaccinated state V to count those who have been vaccinated prior to their first SARS-CoV-2 infection. In terms of protection against infection and severe disease, this state is equivalent to the R state counting those with immunity from prior infection.

To allow for different SARS-CoV-2 variants to have different transmission rates and different IFRs, the compartments E and I are further broken down by variant i, where i indexes the original variant, and the Alpha, Delta, and Omicron variants.

The rate at which agents leave the E_i compartment for both the normal and more transmissible variants is σ and the rate at which agents leave the I_i compartments for all variants is γ. We also include compartments E_i and I_i corresponding to those experiencing breakthrough Omicron infections. These individuals are modeled as having immunity to previous variants but not to Omicron. The purpose of these additional states is to allow the IFR for breakthrough infections to differ from that of other infections.

With these assumptions, the mean generation time for the model is then $1/\sigma + 1/\gamma$. We set this generation time in line with estimates from the CDC.[12] As mentioned above, this generation time sets the time scale of the epidemic implied by the model.

III.C. The Model of Behavior and Disease Transmission

We use an ad hoc model of the impact of behavior on transmission rates. Specifically, the reduced form for the behavioral response of the transmission rate to the level of daily deaths is given by

$$\beta_i(t) = \bar{\beta}_i \exp\left(-\kappa(t)\frac{dD(t)}{dt} + \psi(t)\right)$$

where the parameters $\bar{\beta}_i$ control the inherent transmissibility of the original and subsequent variants of SARS-CoV-2, the parameter $\psi(t)$ is used to introduce seasonality in transmission, and $\kappa(t)$ is the semi-elasticity of transmission with respect to the level of daily deaths. Thus, public and private behavior having an impact on transmission is assumed to respond only to the current level of daily deaths.

Five comments regarding this model of behavior are in order.

First, we have assumed that behavior reacts to the current level of daily deaths. As described in Atkeson (2021b), this form of behavior serves to regulate the effective reproduction number and drive it down to one in the initial phase of the epidemic and then keep it close to one for the remaining course of the epidemic. More specifically, such behavior regulates the model-implied growth rate of cumulative deaths to remain roughly

12. See CDC, "COVID-19 Pandemic Planning Scenarios," https://archive.cdc.gov/www_cdc_gov/coronavirus/2019-ncov/hcp/planning-scenarios.html. On that web page, the CDC notes a mean time of approximately six days between symptom onset in one person to symptom onset in another person infected by that individual.

constant over time. We argue throughout this paper that this outcome of roughly linear growth of cumulative COVID-19 deaths is one of the most striking features of the data on COVID-19 deaths, not only in the United States but around much of the world. While this outcome might be predicted by economic theory, it is not universally observed across epidemics. For example, as we discuss in the online appendix, mitigating behavior seems to have taken a different and more persistent form in the recent mpox epidemic. Thus, it is not clear that behavior will take the same form in the next epidemic.

Our second comment concerns the role of the dual assumptions that behavior responds to the daily death rate and not the level of infections and that, due to the presence of the H compartment, daily deaths are essentially a distributed lag of past levels of $I(t)$. As discussed in the appendix of Atkeson (2021b), these assumptions appear to be remarkably successful in allowing the model to match the size of the waves of COVID-19 deaths with each new variant over the past four years. Models in which behavior reacts to the level of infections directly or that do not include this lag have difficulties in matching the size of these waves as, in these cases, behavior is too successful at keeping the effective reproduction number close to one. That is, mitigating behavior reacts so quickly to changes in the level of infections that waves are cut off.[13]

Third, we see the introduction of new variants as exogenous shocks to transmission rates that allow us to identify the strength and timing of the behavioral response of the model. We thus take the observation that the model can match the size and shape of the waves of deaths associated with the introduction of the Alpha, Delta, and Omicron variants as validation of the parameter choices governing the behavioral response in the model, including the delay induced by the H compartment. Moreover, we take from the Omicron wave in which new infections spiked much higher than in previous waves and much higher than deaths did as validation of the assumption that behavior responds to deaths and not infections.

Fourth, the waves of COVID-19 deaths appear to have a seasonal pattern, with summertime lows, which we match with our seasonal factor $\psi(t)$ chosen to follow a sine wave.

Fifth, ideally, one would want to build a model in which agents are fully rational and make decisions about their mitigation behavior, particularly for understanding behavioral responses in the counterfactuals that we

13. On this point, see Droste and Stock (2021) and Atkeson, Kopecky, and Zha (2021).

consider. To build such a model, however, one must take a stand on what agents believe about the risks that they face from the disease, and this can be hard to do in real time. Moreover, it might also be difficult to incorporate the delayed responses of behavior that appear to be critical in reproducing the dynamics of the epidemic that we have observed. We leave these challenges to future research.

Fitting this model to the data has been an ongoing project starting with a first version in Atkeson (2021a). The goal has been to explore whether one could account for the dynamics of the COVID-19 epidemic with a simple model with a stable formulation of behavior. To that end, in previous work and in this model, we find that the strength of the response of public and private behavior having an impact on transmission to the level of daily deaths as indexed by $\kappa(t)$ appears to have relaxed in the late fall of 2020 and remained consistent since then. Specifically, we choose an initially high value of $\kappa(t)$ for the period February 15, 2020 until November 2020, and then $\kappa(t)$ declines to a new level equal to 35 percent of its initial value. We refer to this apparent relaxation of behavior in the face of the level of daily COVID-19 deaths as "fatigue." We find that this onetime change in behavior in our model is required for the model to match the height of the waves of COVID-19 in late 2020 and beyond.[14] This formulation of behavior was chosen early on in this modeling process starting with the first version of this model in February 2021 and has been kept constant since that time.

III.D. Key Parameters

We set the IFRs for the SARS-CoV-2 variants prior to Omicron to be a declining function of time. As discussed above, the serology data estimates for the percentage of the population infected as of the end of 2020 and the data on cumulative deaths at that time imply an IFR of 1 percent for 2020.

We use that value for the IFR for 2020. The corresponding IFR implied by the serology and deaths data for 2021 for the period prior to Omicron is 0.5 percent.

To match the big jump in infections with the first Omicron wave with an increase in deaths that is modest in comparison to what would have happened if Omicron was as deadly as prior variants, we use a lower IFR of 0.15 percent for those infected with Omicron out of the S compartment. We also allow Omicron to infect those in the R compartment (those with

14. Andersson and others (2021) and De Gaetano and others (2023) argue that the impending arrival of effective vaccines may have caused such a relaxation of behavior.

protection from prior infection or vaccination) with a very low IFR. We refer to such infections as *breakthrough infections*.

Having chosen these parameters, we choose the parameters for inherent transmissibility $\bar{\beta}_\iota$ to match the dynamics of the waves of deaths associated with each of them. As described in the online appendix, these parameters imply a relative transmissibility across variants indexed by the ratio of these parameters that is in line with established estimates.

In modeling the transmissibility of Omicron, one must set two parameters—the constant $\bar{\beta}_\iota$ reflecting its inherent transmissibility and a parameter governing the probability that a vaccinated or recovered individual suffers a breakthrough infection. These two parameters combine to give Omicron a growth advantage over Delta. There is considerable uncertainty regarding the relative importance of these two parameters. We choose them to match data from South Africa that Omicron had a growth advantage of a factor of three relative to Delta in a population with 85 percent protected by prior immunity as well as our serology and deaths data in that first Omicron wave.[15] We find that our model's implications for the first wave of Omicron deaths are largely invariant to the particular choice of $\bar{\beta}_\iota$ for Omicron. What does vary as this parameter is varied (and the probability of a breakthrough infection modified to keep the growth advantage of Omicron over Delta at three times) is the size of the wave of initial Omicron infections. We have chosen a pair of parameters to match this growth advantage for Omicron and this wave of infections as indicated in the serology data.

To model the impact of vaccines, we set the rate at which susceptible agents are moved from the S compartment directly to the V compartment equal to $\lambda(t) = 0.0065$ starting on January 1, 2021, and zero before that date. Vaccines are administered at this rate for the first 185 days of 2021. The rate of vaccination then drops to $\lambda(t) = 0.0065/5$ until the end of 2022 and then $\lambda(t)$ is set to zero after that. In the model, the V compartment is equivalent to the R compartment and is simply used to count vaccinations prior to infection.

In our model, agents in compartment $V(t)$ enjoy full protection from infection by the Alpha and Delta variants and substantial protection against death from Omicron in the same way as agents with prior infection (in the R compartment). Thus, we regard the number of agents in this compartment

15. See, for example, Raquel Viana, Sikhulile Moyo, Daniel G. Amoako, Houriiyah Tegally, Cathrine Scheepers, Christian L. Althaus, and others, "Rapid Epidemic Expansion of the SARS-CoV-2 Omicron Variant in Southern Africa," *Nature* 603 (2022): 679–86.

as representing the population that is both vaccinated prior to a first SARS-CoV-2 infection and that gained protection from that vaccination. To model that vaccines are not 100 percent effective, we assume that the portion of those who arrive in the V compartment is 75 percent of the total vaccinated. Thus, when we compare the model implications for $V(t)$ to the measures from the serology data on those vaccinated but not infected in figure 4, we plot $V(t)/0.75$ as a measure of the total population vaccinated.

We assume that agents flow out of the V and R compartments back to the S compartment and thus become susceptible again to severe disease at a rate corresponding to expected duration of protection against severe disease of three years. Because Omicron can also infect those in the R and V compartments with breakthrough infections (but with a much lower IFR), our model allows protection against reinfection to wane much faster than protection against severe disease. It is this second process that largely accounts in the model for the long tail of COVID-19 deaths that we see over the past two years. Both estimates for the speed of waning are subject to considerable uncertainty.

IV. Main Model Results: Four Lessons

We now use the model to conduct counterfactual experiments to explore the impact of behavior and vaccines on cumulative mortality from COVID-19 in the United States over the past four years. We focus on drawing four lessons from the model.

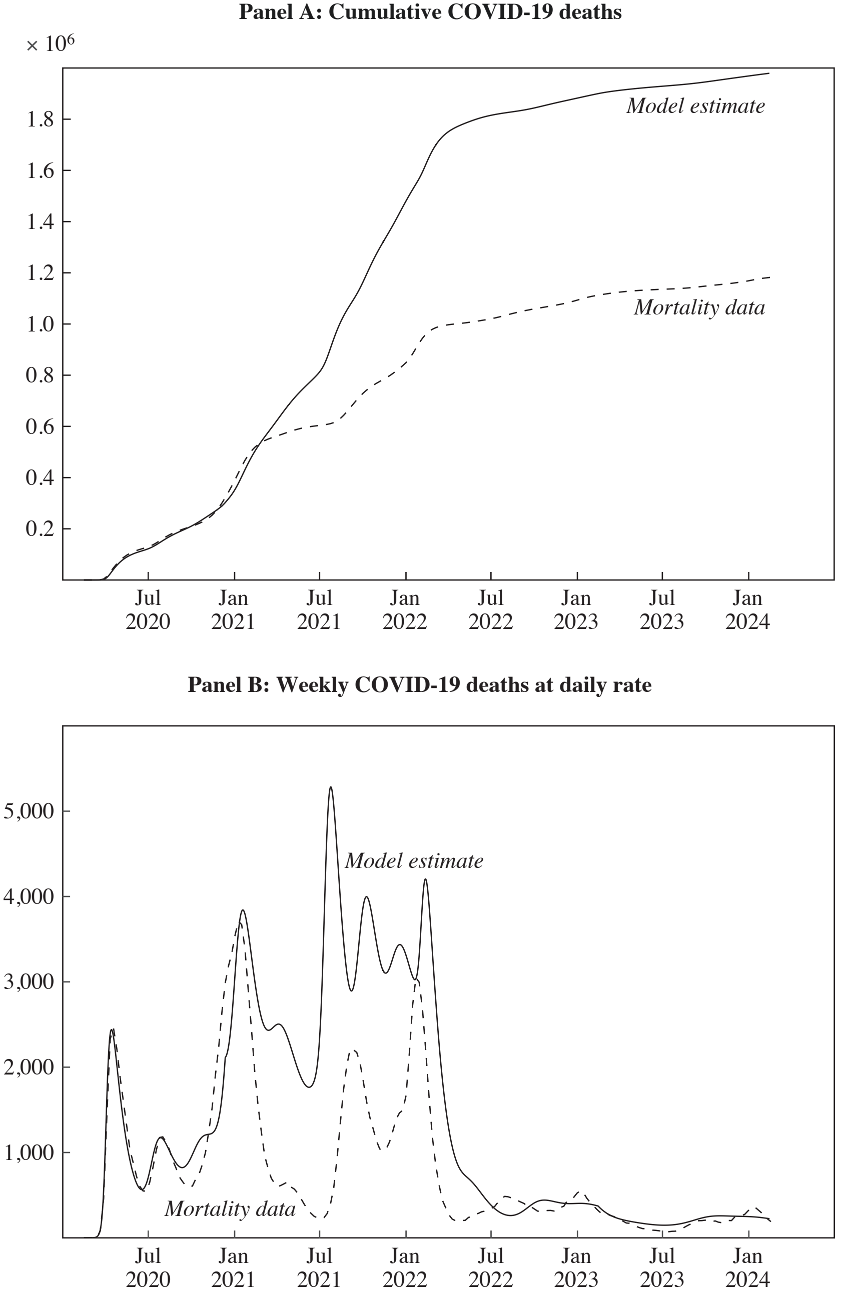
IV.A. Lesson 1: Behavior and Vaccines Together

As discussed above, we show our model's baseline implications for the dynamics of COVID-19 deaths, infections, and vaccinations in figures 2 and 4. We show the model's baseline implications for cumulative COVID-19 mortality over the four-year period from February 15, 2020 to February 15, 2024 in the first row of table 1 above.

We show the model implications for COVID-19 deaths with the baseline parameters governing behavior but with no vaccines in figure 5 and in the second row of table 1. As indicated in the second row of table 1, the model implies that absent vaccines, the cumulative death toll over the past four years would have been 1,979,000. That is, our model implies that, given baseline behavior, vaccines saved 799,000 lives. We take this as the headline result of this paper.

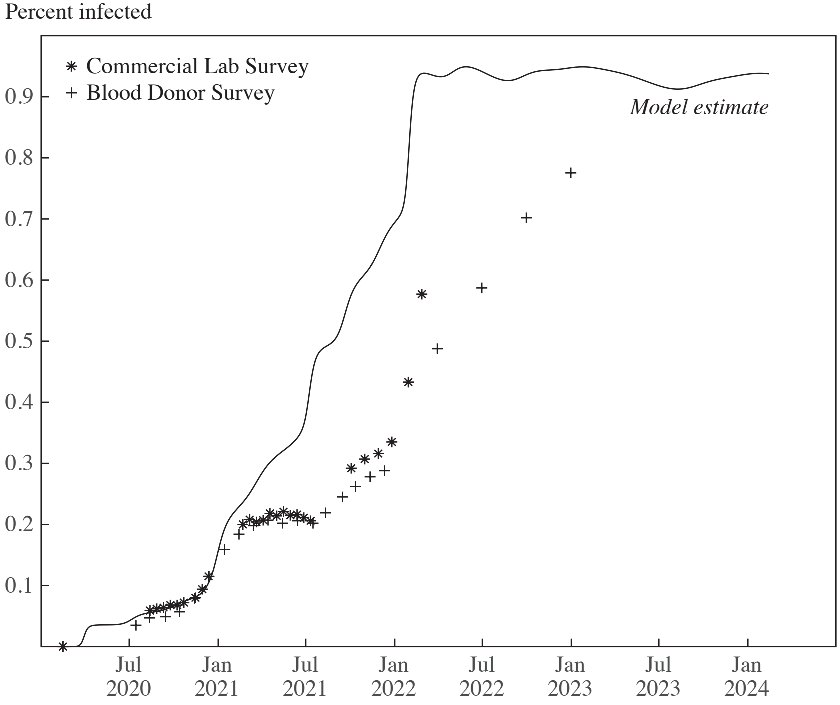
We see in figure 5 that most of these additional deaths would have occurred in 2021. After the first big Omicron wave in early 2022, the model

Figure 5. Baseline Model Behavior but No Vaccines



Source: CDC COVID Data Tracker and authors' calculations.

Figure 6. Dynamics of Infections in the Absence of Vaccines



Source: CDC and authors' calculations.

implications for COVID-19 deaths with and without vaccines are nearly the same. This is because, in the absence of vaccines, the model implies that close to 95 percent of the population would have experienced their first SARS-CoV-2 infection by the end of that first Omicron wave and thus the level of population protection against severe disease after that point would have been similar with or without vaccines. This prediction of our model for the dynamics of infections in the absence of vaccines is shown in figure 6.

In our model, we assume that vaccination reduced the IFR from first infection with SARS-CoV-2 in 2021 from 0.005 to 0.0013 (or 25 percent of the IFR for the naive unvaccinated).[16] Thus, to understand our counterfactual

16. Recall that we assume that 25 percent of those who receive a vaccine do not end up with protection, while the other 75 percent gain complete protection until either their immunity wanes or they suffer a breakthrough infection with Omicron.

estimate of the COVID-19 death toll in the absence of vaccines, imagine that this 68 percent of the population had instead been infected without the protection of vaccines and had, as a result, suffered the full IFR of 0.005 rather than 0.0013. Had this occurred, the counterfactual death toll from COVID-19 in the absence of vaccines would have been 847,000 higher than the baseline with vaccines.[17] Our full model delivers a slightly lower estimate of lives saved due to the arrival of Omicron in late 2021, which had a lower IFR than prior variants, and the assumption that the protection against severe disease offered by vaccines (and prior infection) wanes over time. But one can clearly see from this calculation the simple logic underlying our estimate of the impact of behavior and vaccines on cumulative mortality from COVID-19 in the United States.

These blood serology data highlight how important the interaction of behavior change and vaccine development and deployment were in saving lives. Had SARS-CoV-2 swept through the US population in an unmitigated epidemic, it is likely that the overwhelming majority of the US population would have been infected by early fall of 2020, leaving much less room for people to benefit from being vaccinated prior to their first infection.

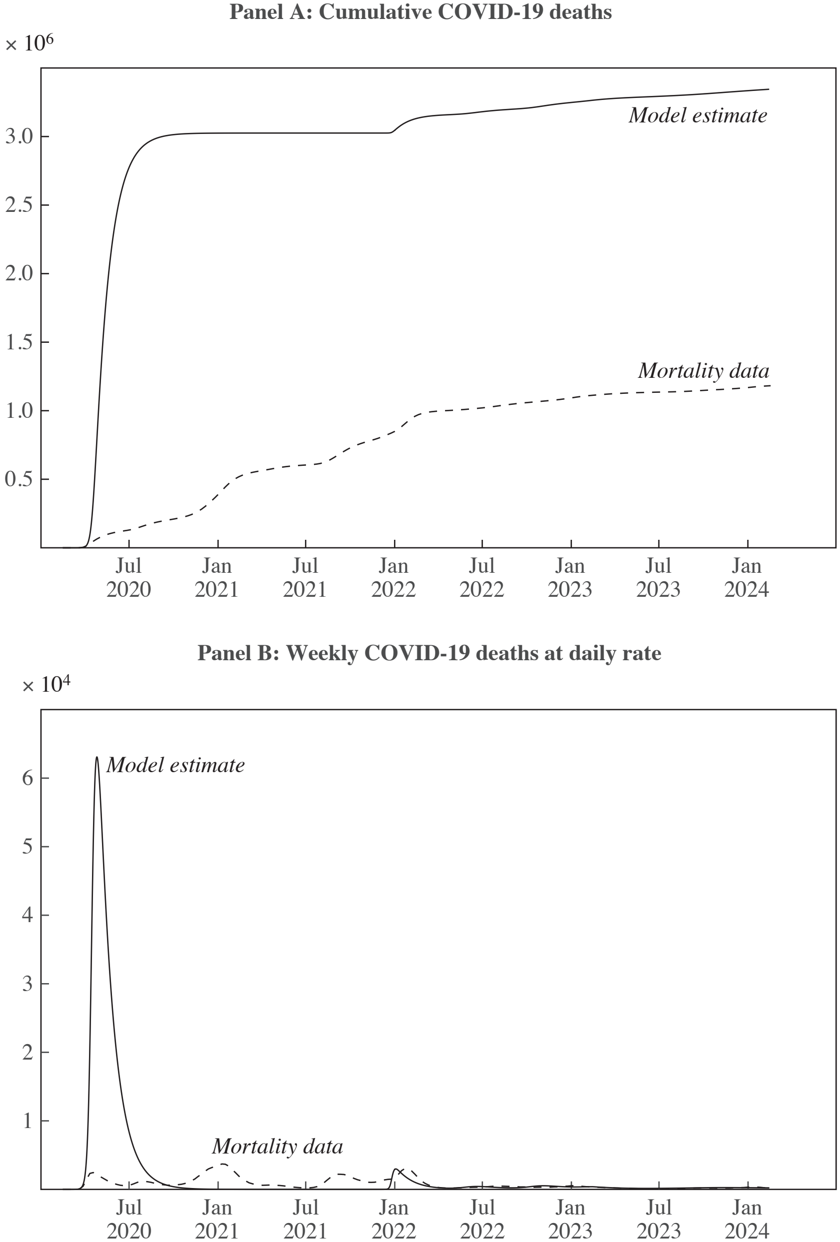
We illustrate this point by simulating the model with the behavioral parameter $\kappa(t) = 0$. As shown in figure 7, in this simulation, the vast majority of the US population gets infected by the late summer of 2020.[18] We report the implied cumulative death toll in the third row of table 1. Here we see an extraordinary model-implied death toll, consistent with an IFR of 1 percent applied to nearly the entire US population in 2020 together with subsequent deaths in later years due to waning immunity.

Clearly, any estimate of lives saved depends on the assumed counterfactual. What impact should this have on our thinking about the next pandemic? From an ex ante perspective as of March 2020, the premises on which our ex post estimation is based would have been hard to predict. Was it going to be possible to delay transmission for the time required to develop and deliver effective vaccines? If vaccines had taken much longer to arrive or had offered less protection against severe disease, would the whole exercise of slowing transmission have been a wasted effort?

17. The calculation is $0.68 \times 0.75 \times 0.005 \times 332{,}000{,}000$ where the last term is the US population.

18. Our results for the cumulative mortality of an unmitigated epidemic during 2020 are worse than those in Ferguson and others (2020) in part because our estimate of the basic reproduction number of the original variant is higher (we assume 3 while they assumed 2.5) and thus an unmitigated epidemic infects more of the population and in part because our estimated IFR at the start of the epidemic is slightly higher (1 percent versus 0.9 percent).

Figure 7. Model with No Behavioral Response



Source: CDC COVID Data Tracker and authors' calculations.

Based on these simulations comparing the death toll with baseline behavior and no vaccines to that with no mitigating behavior, we argue, in short, no—as such, mitigation efforts would have still helped to reduce strain on a severely overburdened health care system and bought critical time to learn how to better care for patients with severe disease even in the absence of vaccines. Such considerations are important to bear in mind when considering which behavioral interventions should be adopted. Without behavioral responses to the epidemic, an unmitigated epidemic would have been much more severe than even our counterfactual with behavior but without vaccines.

IV.B. Lesson 2: Strength and Duration of the Behavioral Response Was a Surprise

We argue now that the success in slowing the spread of COVID-19 during 2020 and 2021 evident in the serology data came as a surprise relative to both historical experience with pandemic influenza and model-based estimates of the impact of mitigation measures on transmission based on that historical evidence.

In many ways, pandemic influenza was the closest historical and epidemiological parallel to the COVID-19 epidemic. Both diseases are fast-moving respiratory diseases with potentially high IFRs. The case of the 1918–1919 "Spanish Flu" epidemic was viewed as particularly relevant, but the epidemics of 1957, 1968, and 2009 also served as examples.

The risk of a new pandemic influenza has been viewed as a substantial threat for a long time. See, for example, the disease and economic scenarios laid out by the President's Council of Economic Advisers in September 2019 (CEA 2019), which foresaw the potential for hundreds of thousands of deaths and trillions of dollars of economic disruption from a pandemic influenza.

In response to this threat from pandemic influenza, epidemiologists have invested considerable effort into studying historical experiences and modeling the impact of various mitigation options on influenza transmission.[19]

19. For examples of studies of transmission during the 1918–1919 pandemic influenza, see Mills, Robins, and Lipsitch (2004), Fraser and others (2011), and Eggo, Cauchemez, and Ferguson (2011). For studies of the impact of mitigation on transmission during the 1918–1919 pandemic influenza, see, for example, Bootsma and Ferguson (2007), Hatchett, Mecher, and Lipsitch (2007), Correia, Luck, and Verner (2022), and Velde (2022).

Of particular interest in this regard is figure 1 in Hollingsworth and others (2011), which shows the duration (in weeks) and effectiveness (in terms of percentage reduction in transmission rates) of historical interventions to slow the spread of the 1918–1919 influenza and SARS-CoV-1. That figure estimates that interventions in the 1918–1919 influenza pandemic reduced transmission rates by less than 50 percent in all cases and much less than that amount in many cases. Moreover, these interventions were sustained for less than fifteen weeks. As shown in that figure, mitigation efforts for SARS-CoV-1 were estimated to be much more effective, but these were also sustained for less than fifteen weeks. In comparison, with COVID-19, we see from the serology data that efforts to slow disease spread substantially had an impact for many months through late 2021.

Prominent studies of the possibilities for using public health interventions to contain a new influenza strain at its source include Longini and others (2005) and Ferguson and others (2005). Prominent modeling studies of the use of broader public health measures, including school closures and social distancing, to slow the spread of a pandemic influenza that broke through efforts to contain it at the source include Ferguson and others (2006) and Germann and others (2006). Universally, these studies predict short periods of very rapid spread of disease even in the modeled presence of intense public health efforts to slow disease spread so that available flu vaccines can be administered.

Particularly telling in this regard is the caption of figure 4 in Ferguson and others (2006, 451) that notes, regarding the timing of administration of vaccines, these vaccines would have “almost no effect” if started after 120 days after the first worldwide case.[20] This conclusion is clearly too pessimistic about the possibility of controlling the spread of a respiratory pathogen through behavioral mitigation, as COVID-19 vaccines still had a major benefit despite arriving more than a year after the first worldwide case. The COVID-19 pandemic fundamentally changed our conception of what is possible with respect to behavioral mitigation.

This contrast between the anticipated and observed impact of behavioral change on slowing the transmission of COVID-19 is even more remarkable given that the original strain of SARS-CoV-2 was more contagious than a pandemic influenza strain was expected to be, had the ability to spread prior to the onset (or in the absence) of symptoms, and ultimately

20. See also figure 2 in Germann and others (2006).

generated new variants with substantially increased transmissibility. The cards were stacked against us, even relative to the modeled scenarios for pandemic influenza that served as our basis for our earliest understanding of SARS-CoV-2.[21]

To see this point, consider the scenarios for pandemic influenza expected by modelers as laid out in Meltzer and others (2015). Table 1 in that paper lays out the range of scenarios for transmissibility and clinical severity of potential new pandemic influenzas typically considered, and figure 1 in that paper places historical pandemics in this space of transmissibility and clinical severity. The original strain of SARS-CoV-2 had higher transmissibility than the worst-case scenario and was near to the worst-case scenario in terms of its clinical severity.

The fact that SARS-CoV-2 could be transmitted prior to showing symptoms made epidemiologists (including ourselves) pessimistic that its spread could be effectively controlled. As described in Fraser and others (2004, 6146), “the success of . . . control measures is determined as much by the proportion of transmission occurring prior to the onset of overt clinical symptoms (or via asymptomatic infection) as the inherent transmissibility of the etiological agent (measured by the reproductive number R_0).” Likewise, early in the COVID-19 epidemic, Hellewell and others (2020) pointed to the pre- and asymptomatic transmission of COVID-19 as a reason to be pessimistic about our ability to contain its spread.

These features of COVID-19, together with the hazy prospects as of March 2020 for developing an effective vaccine in time to be useful, meant that despite all the planning for a pandemic influenza the set of actionable targeted mitigation policies available to slow the spread of COVID-19 in a cost-effective manner was very small. In fact, in an early and highly cited article from March 9, 2020, giving broad outlines of options for mitigating the coming pandemic, Anderson and others (2020, 934, emphasis added) remarked that “it is easy to suggest a 60 percent reduction in transmission will do it or quarantining within 1 day from symptom onset will control transmission, *but it is unclear what communication strategies or social distancing actions individuals and governments must put in place to achieve these desired outcomes*.”

We argue that one of the main lessons of our experience with COVID-19 is there are far greater possibilities for slowing transmission of a deadly respiratory virus than previously thought. Given that new knowledge, we

21. See, for example, Davies and others (2020).

should work urgently to determine how to achieve similar behavioral mitigation in the next pandemic but at far lower cost.

IV.C. Lesson 3: Behavior and State-Level Outcomes

There has been great interest in comparing the impact of COVID-19 across states of the United States in the press and in some academic work.[22] Certainly the outcomes for cumulative mortality for COVID-19 vary widely across the states of the United States. What accounts for these differences? We address this question in greater detail in section B of our online appendix.

Here we make the argument that, relative to the historical and modeling benchmarks for pandemic influenza discussed above, residents of all fifty states made surprisingly strong and lasting efforts to slow the spread of SARS-CoV-2 so that vaccines came in time to save a considerable number of lives.

To illustrate this point, in online appendix figure B.6, taken from Chitwood and others (2022), we show the dynamics of the effective reproduction number for SARS-CoV-2 for each of the fifty states of the United States. In this figure, we observe that behavior in all fifty states changed rapidly and dramatically so as to drive the effective reproduction number of COVID-19 in the state down to one very early on in the epidemic. Moreover, this behavior was sufficiently sustained to keep this effective reproduction number close to one throughout 2020. Atkeson, Kopecky, and Zha (2024) find similar results for both US states and many countries.

As we have discussed above, if the effective reproduction number of a disease remains close to one, then the growth rate of current infections is close to zero. Equivalently, the growth rate of cumulative infections and deaths is then roughly constant. This is precisely the dynamics we observe in cumulative COVID-19 mortality at the state level.

To illustrate this point, in online appendix figure B.7, we show the dynamics of cumulative COVID-19 deaths as an age-adjusted death rate per 100,000 of the population for selected states. In the left panel of this figure, we show the dynamics of cumulative COVID-19 deaths for California, Florida, New York (excluding New York City), and Texas. We see that New York State had a very rapid growth of cumulative deaths in the initial phase of the pandemic and then settled into a lower growth rate. Texas had

22. See, for example, Barro (2022), Bollyky and others (2023), and Kerpen, Moore, and Mulligan (2022).

a high growth rate of cumulative deaths throughout the first two years of the pandemic. Given the rhetoric surrounding this topic, we find it striking how similar the age-adjusted outcomes for COVID-19 deaths have been for California and Florida over the past four years.

In the right panel of online appendix figure B.7, we show the dynamics of cumulative COVID-19 deaths as an age-adjusted death rate per 100,000 of the population for New York City and seven other states representing extreme high and low mortality outcomes across states. With the exception of New York City, we see largely linear growth in cumulative deaths over the first two years of the COVID-19 epidemic for all of these locations. As evident in the figure, New York City suffered exceptionally rapid initial growth of cumulative COVID-19 deaths in the first wave of the epidemic, likely due to the surprise introduction of a large number of hidden cases from Europe in early 2020.

For further evidence of this commonality of responses across US states, in online appendix figure B.8, we show estimates from the Commercial Lab and Blood Donor serology surveys of cumulative infections and combined seroprevalence for the fifty states of the United States. While these surveys show considerable variation in the estimated percentage infected across states, we see in this figure that all of the states followed similar dynamics of slow growth in infections in the first two years of the pandemic and rapid deployment of vaccines in the first half of 2021.[23] We discuss these state-level serology data in greater detail in online appendix section B.

Based on this evidence, we argue that the most important feature of the outcomes across US states (and even countries around the world) is how much they have in common relative to outcomes that were expected given prior epidemiological modeling of and past experiences with pandemic influenza. To a large extent, residents of every state in the United States outside of New York City reacted very strongly to COVID-19 very early on and took significant actions to slow its spread all through 2020 and 2021. We regard the observation that this could be done and done nearly universally across different states of the United States, as a great surprise.

To expand further on this point, observe that the model-based forecast in Ferguson and others (2020) for peak deaths with unmitigated spread of COVID-19 was over sixteen deaths per day per 100,000 population (implying over 50,000 deaths per day in the United States as a whole) with

23. Chitwood and others (2022) argue that the serology data underestimate the true portion of the population ever infected for a variety of reasons. This paper presents alternative estimates of the state-level portion of the population infected through 2020 in its figure 7.

75 percent of the population being infected by late summer of 2020. This forecast was not out of line with what was experienced in locations that did little to mitigate the spread of SARS-CoV-2. For example, we note that seroprevalence studies in Manaus, Brazil indicated an attack rate of 75 percent in the first wave of the pandemic (Buss and others 2021). We see nothing like this rapid spread of COVID-19 in the serology data across US states.

In March and April 2020, New York City experienced the worst wave of COVID-19 cases and mortality of anywhere in the United States over the past four years. Its peak weekly mortality rate was sixty per 100,000 population (less than ten per 100,000 per day)—in the range of one-half that predicted in Ferguson and others (2020) for peak deaths with unmitigated spread. Seroprevalence estimates for New York City indicate up to 20 percent of that population of 8 million people was infected in the first wave in the spring of 2020 (Stadlbauer and others 2021).

We illustrate the extent to which the first wave of COVID-19 deaths in New York City was an outlier in online appendix figure B.9. In that figure, we show the dynamics of weekly COVID-19 deaths for the fifty states at an age-adjusted rate per 100,000 of population. As is clear from the figure, the first wave of COVID-19 deaths in New York City was much larger than any other wave experienced in any state in the United States. That is, the response to flatten the curve and dramatically slow the transmission of COVID-19 was universal across the fifty states of the United States.

We now turn to our fourth lesson regarding the prospects for a similar behavioral response next time.

IV.D. Lesson 4: Unclear If Behavior Will Be the Same Next Time

From our perspective, the success of this sustained and fairly uniform behavioral response to slow transmission for this length of time to allow for the deployment of vaccines and improved medical care is perhaps the biggest surprise of the COVID-19 pandemic. Clearly, a strategy of slowing transmission for eight to fifteen months as needed to develop and deliver an effective vaccine is based on the premise that people can be persuaded to go along with that plan. To an extent that seems well outside historical experience with pandemic influenza and predictions based on that historical experience, Americans did go along with that plan, with or without mandates from state governments.

Many economists, one of us included, have argued ex post that this pattern of adjusting behavior to keep the growth rate of new infections and deaths relatively close to zero, observed nearly universally in the United

States and across many countries, is precisely the response that economic theory would predict.[24]

But this argument then raises the puzzle of why did we not see a quantitatively similar response to pandemic influenzas, in particular the 1918 Spanish Flu? And comparison of these different outcomes for pandemic influenza and COVID-19 raises the question of which behavioral response should we expect to see in the next pandemic? Will it be a short, sharp wave as for COVID-19 in New York City in March and April 2020 and in most cities for which we have data from 1918? Or will it be a long, drawn-out affair as for COVID-19 in the rest of the United States? The answer to this question will have a big impact on the range of mitigation strategies available in the face of the next pandemic and is a great challenge in epidemiological modeling (Funk and others 2015).

The world has already experienced an outbreak of another emerging pathogen. Starting in May 2022, mpox, formerly known as monkeypox, began to spread rapidly primarily through sexual contact between men, with this spread being particularly alarming since it showed up in a large number of countries in a short period of time. Mpox is an example of a known pathogen endemic to a relatively small area (in West Africa) suddenly spreading rapidly well outside that region.

Simple examination of the exponential growth of cases in the United States between May and August 2022 indicated that this disease had the potential to spread quite broadly, at least within a subset of the US population. Instead, the number of cases began to die out rapidly in late August, and new cases in the United States have been held at a low level throughout 2023. What explains this path of this epidemic? It appears that through a combination of a sustained change in private sexual behavior and the targeted application of vaccines, it was possible to dramatically reduce the number of cumulative cases relative to what would be predicted for an unmitigated epidemic. In other words, the behavioral response to mpox appears to be a remarkable success.

Zhang and others (2024) quantify the impact of behavior and vaccines on the spread of mpox with an epidemiological model using data on the mpox outbreak in the United Kingdom, which exhibited an epidemic curve similar to that in the United States. These authors argue that changes

24. See Atkeson (2021b, 2023a) and Atkeson, Kopecky, and Zha (2024). See Gans (2022) for a broader survey of the economics papers on this topic.

in behavior and vaccination together played an important role in shaping this epidemic.

We find it interesting to note, however, that their accounting of the impact of behavior and vaccines on the trajectory of this epidemic is quite different from our accounting of the impact of these factors on the trajectory of the COVID-19 epidemic. In particular, they find that the response of behavior (in terms of men reducing the number of their sexual partners) was strong and persistent enough to drive the effective reproduction number of mpox below one on a sufficiently sustained basis to drive the number of new cases to a very low level. This never happened with COVID-19. They then estimate that the use of pre-exposure vaccines for susceptible men limited the threat of resurgence.

The estimated combined impact of these interventions was then very substantial in limiting the size of the outbreak: the United Kingdom had 3,250 observed cases over the study period relative to an estimated final size of an uncontrolled epidemic of 169,400 cases. We see these estimates, together with the discussion in Daskalakis, Romanik, and Jha (2024), as driving home the message that targeted interventions in combination with vaccination can have a powerful impact on outcomes of an epidemic.

Another factor to consider going forward is the extent to which our experience with COVID-19 will shape reactions to new epidemics going forward for decades to come. Will the public be more skeptical of public health warnings about new pathogens? Or will our collective experience with significant mortality from an infectious disease outbreak lead us to take future threats more seriously? Addressing such questions seems of first-order importance for research going forward.

V. What Is Needed to Make Mitigation Less Painful Next Time?

The behavioral mitigation measures undertaken during the COVID-19 pandemic helped to save many thousands of lives, but they came at a high social and economic cost. Uncertainty about key features of COVID-19 and about the human behaviors that had an impact on its spread forced us to take stronger, more widespread, and longer-lasting behavioral mitigation measures than might have been necessary in a more information-rich setting. Likewise, individuals largely lacked the tools they needed to make informed assessments about their risk of becoming infected or transmitting disease. For example, widespread and cheaply available diagnostic tests—along with clear guidance on how to report and interpret them—could have helped alleviate the need for general physical distancing measures like school and workplace closures that lasted for many months into the pandemic.[25]

The next pandemic may look very different from COVID-19, but it will nevertheless be critical to find ways to rapidly reduce our uncertainty about the pathogen's characteristics and the human behaviors that underlie its spread, and likewise to rapidly develop and deploy the tools that will empower individuals to make informed behavioral choices. This will require developing off-the-shelf research protocols for learning about transmission routes, the natural history of infection, and the dynamics of immunity for an emerging pathogen soon after it is first detected. In the meantime, we must also invest in ongoing data collection efforts to provide baseline measurements against which data on an emerging infectious disease can be meaningfully compared. A detailed discussion of the steps needed to effectively prepare for the next pandemic is provided by Lipsitch and others (2023). Here, we outline a few key considerations.

V.A. Assessing Transmission Routes

When an emerging outbreak is detected, a critical first task is to determine the pathogen's routes of transmission. Beyond the most basic information on transmission route (e.g., sexual versus vector-borne versus respiratory, and [if respiratory] droplet versus aerosol versus fomite), it is also important to identify the venues and behaviors that are most conducive to spread. For example, it became evident early in the COVID-19 pandemic that outdoor transmission was far less common than indoor transmission (Bulfone and others 2021) and that singing was a particularly high-risk activity (Hamner and others 2020). Preapproved study designs, backed with funding for rapid deployment, would help to more rapidly clarify how and where the bulk of transmission occurs in the event of an emerging outbreak.

To place these studies in the proper context, we also require detailed studies on interpersonal contact patterns, both at baseline and as they evolve over the course of an outbreak, much like the CoMix study did in the context of COVID-19 (Gimma and others 2022). Such studies recruit representative cohorts and ask questions about their behaviors (e.g., conversational or sexual contacts) that may be relevant to the spread of disease. Mobility data—gathered, for example, using mobile phones—can also be useful (Buckee and others 2020), though such data must be interpreted with

25. See, for example, the discussion in Atkeson and others (2020).

care since the owners of mobile devices or the users of a given app may not be representative of the broader population (Wesolowski and others 2016). Data access and privacy issues should also be proactively addressed well in advance of a public health crisis.

Detailed contact tracing data can be useful for determining the level of risk associated with various types of contact. For respiratory infections, household transmission studies like the Office for National Statistics (ONS) Coronavirus (COVID-19) Infection Survey in the United Kingdom (Pouwels and others 2021) can be helpful for assessing the level of risk associated with close contact. For sexually transmitted infections, partnership surveys can serve the same purpose.[26] The value of such studies can be greatly enhanced by collecting pathogen genomic information, allowing researchers to distinguish direct within-household (or within-partnership) transmission from new introductions from the community.

The production and distribution of nonpharmaceutical interventions (NPIs) should be rapidly scaled up in the event of an emerging outbreak. In the early stages of an outbreak, plausible effectiveness should be enough to justify the use of sufficiently low-impact NPIs—for example, plausible effectiveness would justify the widespread use of masks against the early spread of SARS-CoV-2 or condoms to prevent transmission of mpox, even in the absence of direct studies assessing the efficacy of those interventions for those specific pathogens. In tandem, the effectiveness of these NPIs should be continuously monitored so that their use can be founded on more direct, pathogen-specific evidence or, if no effectiveness is found, their use can be phased out.

V.B. Describing the Course of Infection

Once infection occurs, it is critical to understand the risk of various health outcomes. Key statistics like the IFR are subject to bias that can affect early estimates in both directions: early in an epidemic, the most severe cases are the ones that are most likely to be detected, thus skewing the IFR upward; yet, if the epidemic is spreading rapidly, a simple division of mortality by cumulative prevalence can skew the IFR downward, since recently infected individuals have not yet had time for their cases to worsen. This underscores the need for principled studies to track the range, timing, and probability of potential health outcomes in an emerging epidemic. An understanding of the IFR and related risks of various health outcomes helps to set the appropriate level of behavioral response.

26. For example, Ueda and others (2020).

Similarly, it is important to rapidly assess how a person's infectiousness varies over time. Again, household or partnership studies can be helpful, especially when coupled with frequent, quantitative diagnostic testing (e.g., RT-qPCR tests to assess pathogen load) and detailed symptom reporting. A critical piece of information to gather early in an epidemic is how the timing of symptoms relates to infectiousness, as this relationship plays a major role in determining how difficult it is to ultimately control a pathogen's spread (Fraser and others 2004). If infectiousness precedes symptoms, the need to develop and deploy rapid diagnostic tests becomes paramount.

V.C. Tracking Incidence and Immunity

Public health response in the United States is largely coordinated at the state level, which poses major challenges for data sharing and standardization. The need for improved data collection, standardization, and dissemination is a major focus area of the new Center for Forecasting and Outbreak Analytics (CFA) based at the CDC. The CFA has taken many cues from the National Weather Service (George and others 2019), and indeed a digital infrastructure for providing information on current epidemiological conditions and a near-term forecast would go a long way toward informing more targeted behavioral responses in the event of another public health crisis.

Alongside information on disease incidence, well-designed serological studies can be invaluable both for reconstructing what has happened after an outbreak ends (as we have tried to do in this report) and for informing on the dynamics of immunity. It is important to conduct ongoing serological studies so that proper baselines can be set, especially because serological tests can cross-react.[27] Serological studies can inform on the duration of immunity to infection, thus helping individuals to calibrate their behavior to better match their risk of infection.

VI. Conclusion

The behavioral response to COVID-19 in 2020–2022 was highly—and unexpectedly—effective in reducing cumulative COVID-19-related mortality in the United States. We estimate that the combination of behaviorally driven transmission reduction and vaccination resulted in roughly 800,000 lives saved during that time period, in line with other estimates.

27. For example, serological tests for SARS-CoV-1 can turn positive based on exposure to a related common coronavirus; see Patrick and others (2006).

Critically, we see that both of these factors—a strong behavioral response and the relatively fast development of an effective vaccine—were needed to yield a substantial reduction in mortality. Had a vaccine not been developed or had behavior not changed, we anticipate that much of the US population would have received their first immunological exposure to SARS-CoV-2 from infection rather than vaccination, and thus the total mortality from the pandemic would have been much higher.

We had three main goals in writing this report: (1) we sought to provide an evidence-based estimate of the value of behavior change during the COVID-19 pandemic in terms of reduced mortality; (2) we sought to describe a straightforward modeling framework that can be adapted to assess counterfactual scenarios for COVID-19 and for infectious diseases more generally; and (3) we sought to discuss the lessons of the COVID-19 pandemic from three hypothetical perspectives: the ex ante perspective of a public health planner in March 2020, with knowledge of basic parameters of the virus but no certainty about its future evolution; the ex post perspective, where we are today, performing an assessment of how we actually performed given our knowledge of how the pandemic actually unfolded; and the perspective of future public health planners, who will be responsible for responding to new, possibly very different, emerging infectious diseases. We now discuss each of these goals in turn.

There are many ways to estimate lives saved during the COVID-19 pandemic, some relying on sophisticated models of transmission and immunological dynamics. We pursue a simpler tack, estimating the total mortality in the scenarios with no behavioral change prior to rollout of a vaccine in January 2021 and with behavior change but no vaccine. Based on serology data, we estimate that less than 20 percent of the US population—and a substantially smaller fraction of individuals over age 65—had been infected with SARS-CoV-2 before the introduction of vaccines. Yet other areas of the world that experienced an impact early on, before an effective behavioral response could be mounted, saw estimated attack rates of up to 75 percent within a short few months, which also aligns with epidemiological models for an unmitigated epidemic with transmissibility similar to the ancestral strains of SARS-CoV-2. As such, we can attribute a mortality reduction in the roughly 55 percent of the population who were able to be vaccinated prior to their first infection to the transmission-reducing behavioral response. Had a successful vaccine not been developed, however, it is unclear whether behavioral response would have had a substantial impact on cumulative mortality through the present day, since immune evasion and increasingly contagious variants of the virus have rendered herd immunity moot.

By introducing a modeling framework, we were able to compare more nuanced counterfactual scenarios and to better separate the impact of behavior from that of vaccination. The framework we discuss here is completely standard, perhaps with the exception of the form of the behavior term, which reduces the transmissibility parameter β_j proportionally to an exponentially decaying function of a parameter κ_j that captures the strength of the behavioral response relative to some disease metric (e.g., total infections or the rate of increase in mortality). It is possible to compare many counterfactual scenarios using this framework, but the main takeaway is that behavioral transmission reduction and vaccination have a powerful positive synergy, where the timing of both is paramount—that is, early behavior change, coupled with the rapid development of an effective vaccine, can pay dividends in reduced mortality.

The ex ante perspective of the public health planner in March 2020 is one lacking in many critical details about the pandemic's ultimate course, and yet it is perhaps the most informative perspective to consider when assessing the best course of action in future pandemics. Early modeling work during the COVID-19 pandemic, including our own, anticipated that SARS-CoV-2 would become endemic (Kissler and others 2020; Shaman and Galanti 2020; Murray and Piot 2021) but failed to anticipate both the ratcheting transmissibility of the virus with successive variants and the relatively swift development of an effective vaccine. The ex post perspective is useful for determining what we might have done differently, but this has limited application for future pandemics.

Instead, rather than thinking about what we should have done differently in hindsight from a management perspective and doing that going forward, we should instead ask what types of information we would have wanted during the early days of the pandemic to make more informed, ex post–like decisions and determine how best to put mechanisms in place now to collect that data. For example, key elements of the natural history of infection and the route of infection—such as the frequency of asymptomatic infections, the role of presymptomatic transmission, and the importance of aerosols in transmission—were unclear for far longer than they should have been.

Developing protocols for rapidly identifying cases, charting their course, and determining likely routes of transmission through prospective household and contact surveys, like the ONS Coronavirus Infection Survey in the United Kingdom (Pouwels and others 2021) and the European CoMix Survey (Gimma and others 2022), are critical for future pandemics. Likewise, it is clear that behavior can change spontaneously in response to a perceived infectious threat. It is less important to have an exact model for how behavior changes in response to threat than it is to have a robust framework for measuring the relevant changes in behavior when they actually happen. This will require a robust survey-taking machinery to be rapidly deployed in the event of an emerging pathogen. Such work may be augmented by the development of secure contact tracing technologies, like the ones developed for contact notification during the COVID-19 pandemic. Regardless, we must avoid the trap of "fighting the last pandemic," recognizing that while another coronavirus pandemic could occur within our lifetimes, there are many other threats that should be carefully thought through and incorporated into the data-collecting mechanisms discussed here. That said, the experience with mpox, and the fact that behavioral mitigation measures during the COVID-19 pandemic strongly suppressed the spread of various other pathogens (Koutsakos and others 2021), suggests that behavioral mitigation can be an important tool for addressing a wide range of infectious disease threats.

Our findings are limited by a substantial degree of uncertainty in the actual number of infections that occurred during the pandemic and a lack of reliable data capturing the dynamics of behavioral change during the pandemic. Regarding the lack of behavioral data, it is unclear even what an ideal data set would look like, given that we do not have a solid grasp on what types of interactions are necessary and sufficient for the transmission of a respiratory pathogen.[28] Conversational encounters are often used as a proxy, but the precise dynamics of interpersonal transmission in real settings remain poorly understood. The models we use are intentionally simplified and so gloss over much important variation in baseline risk factors, population structure, and viral attributes that can, and do, have a major impact on transmission patterns. Our goal here is to provide a scaffold to guide thinking about behavior-modulated disease transmission, rather than to faithfully recapitulate the dynamics of a particular outbreak—though we note that, under reasonable assumptions, a fairly faithful recapitulation of those dynamics is possible with a model like the one presented here.

ACKNOWLEDGMENTS We would like to acknowledge many helpful conversations with Ben Moll, Michael Droste, Karen Kopecky, James Stock, and Tao Zha.

28. See, for example, Ferretti and others (2023).

References

Anderson, Roy M., Hans Heesterbeek, Don Klinkenberg, and T. Déirdre Hollingsworth. 2020. "How Will Country-Based Mitigation Measures Influence the Course of the COVID-19 Epidemic?" *The Lancet* 395, no. 10228: 931–34.

Andersson, Ola, Pol Campos-Mercade, Armando N. Meier, and Erik Wengström. 2021. "Anticipation of COVID-19 Vaccines Reduces Willingness to Socially Distance." *Journal of Health Economics* 80: 102530.

Atkeson, Andrew G. 2021a. "A Parsimonious Behavioral SEIR Model of the 2020 COVID Epidemic in the United States and the United Kingdom." Working Paper 28434. Cambridge, Mass.: National Bureau of Economic Research.

Atkeson, Andrew G. 2021b. "Behavior and the Dynamics of Epidemics." Working Paper 28760. Cambridge, Mass.: National Bureau of Economic Research.

Atkeson, Andrew G. 2023a. "COVID-19: Epidemiological Models." *Annual Review of Financial Economics* 15: 7–27.

Atkeson, Andrew G. 2023b. "The Impact of Vaccines and Behavior on U.S. Cumulative Deaths from COVID-19." Working Paper 31525. Cambridge, Mass.: National Bureau of Economic Research.

Atkeson, Andrew G., Michael C. Droste, Michael Mina, and James H. Stock. 2020. "Economic Benefits of COVID-19 Screening Tests." Working Paper 28031. Cambridge, Mass.: National Bureau of Economic Research.

Atkeson, Andrew G., Karen A. Kopecky, and Tao Zha. 2021. "Behavior and the Transmission of COVID-19." *American Economic Association Papers and Proceedings* 111: 356–60.

Atkeson, Andrew G., Karen A. Kopecky, and Tao Zha. 2024. "Four Stylized Facts about COVID-19." *International Economic Review* 65, no. 1: 3–42.

Bajema, Kristina L., Ryan E. Wiegand, Kendra Cuffe, Sadhna V. Patel, Ronaldo Iachan, Travis Lim, and others. 2021. "Estimated SARS-CoV-2 Seroprevalence in the US as of September 2020." *JAMA Internal Medicine* 181, no. 4: 450–60.

Barro, Robert J. 2022. "Vaccination Rates and COVID Outcomes across U.S. States." *Economics and Human Biology* 47: 101201.

Bollyky, Thomas J., Emma Castro, Aleksandr Y. Aravkin, Kayleigh Bhangdia, Jeremy Dalos, Erin N. Hulland, and others. 2023. "Assessing COVID-19 Pandemic Policies and Behaviours and Their Economic and Educational Trade-offs across US States from Jan 1, 2020, to July 31, 2022: An Observational Analysis." *The Lancet* 401, no. 10385: 1341–60.

Bootsma, Martin C. J., and Neil M. Ferguson. 2007. "The Effect of Public Health Measures on the 1918 Influenza Pandemic in U.S. Cities." *Proceedings of the National Academy of Sciences* 104, no. 18: 7588–93.

Buckee, Caroline O., Satchit Balsari, Jennifer Chan, Mercè Crosas, Francesca Dominici, Urs Gasser, and others. 2020. "Aggregated Mobility Data Could Help Fight COVID-19." *Science* 368, no. 6487: 145–46.

Bulfone, Tommaso Celeste, Mohsen Malekinejad, George W. Rutherford, and Nooshin Razani. 2021. "Outdoor Transmission of SARS-CoV-2 and Other

Respiratory Viruses: A Systematic Review." *Journal of Infectious Diseases* 223, no. 4: 550–61.

Buss, Lewis F., Carlos A. Prete Jr., Claudia M. M. Abrahim, Alfredo Mendrone Jr., Tassila Salomon, Cesar de Almeida-Neto, and others. 2021. "Three-Quarters Attack Rate of SARS-CoV-2 in the Brazilian Amazon during a Largely Unmitigated Epidemic." *Science* 371, no. 6526: 288–92.

Chitwood, Melanie H., Marcus Russi, Kenneth Gunasekera, Joshua Havumaki, Fayette Klaassen, Virginia E. Pitzer, and others. 2022. "Reconstructing the Course of the COVID-19 Epidemic over 2020 for US States and Counties: Results of a Bayesian Evidence Synthesis Model." *PLOS Computational Biology* 18, no. 8: e1010465.

Chowell, Gerardo, Lisa Sattenspiel, Shweta Bansal, and Cécile Viboud. 2016. "Mathematical Models to Characterize Early Epidemic Growth: A Review." *Physics of Life Reviews* 18: 66–97.

Correia, Sergio, Stephan Luck, and Emil Verner. 2022. "Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu." *Journal of Economic History* 82, no. 4: 917–57.

Council of Economic Advisers (CEA). 2019. *Mitigating the Impact of Pandemic Influenza through Vaccine Innovation*. Washington: Executive Office of the President.

Daskalakis, Demetre, Nikki Romanik, and Ashish K. Jha. 2024. "Lessons from the Mpox Response." *JAMA* 331, no. 5: 387–88.

Davies, Nicholas G., Adam J. Kucharski, Rosalind M. Eggo, Amy Gimma, and W. John Edmunds. 2020. "Effects of Non-pharmaceutical Interventions on COVID-19 Cases, Deaths, and Demand for Hospital Services in the UK: A Modelling Study." *The Lancet Public Health* 5, no. 7: e375–85.

De Gaetano, Alessandro, Paolo Bajardi, Nicolò Gozzi, Nicola Perra, Daniela Perrotta, and Daniela Paolotti. 2023. "Behavioral Changes Associated with COVID-19 Vaccination: Cross-National Online Survey." *Journal of Medical Internet Research* 25: e47563.

Droste, Michael, and James H. Stock. 2021. "Adapting to the COVID-19 Pandemic." *American Economic Association Papers and Proceedings* 111: 351–55.

Eggo, Rosalind M., Simon Cauchemez, and Neil M. Ferguson. 2011. "Spatial Dynamics of the 1918 Influenza Pandemic in England, Wales and the United States." *Journal of the Royal Society Interface* 8, no. 55: 233–43.

Eksin, Ceyhun, Keith Paarporn, and Joshua S. Weitz. 2019. "Systematic Biases in Disease Forecasting—The Role of Behavior Change." *Epidemics* 27: 96–105.

Ferguson, Neil M. 2007. "Capturing Human Behaviour." *Nature* 446, no. 7137: 733.

Ferguson, Neil M., Derek A. T. Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sopon Iamsirithaworn, and Donald S. Burke. 2005. "Strategies for Containing an Emerging Influenza Pandemic in Southeast Asia." *Nature* 437, no. 7056: 209–14.

Ferguson, Neil M., Derek A. T. Cummings, Christophe Fraser, James C. Cajka, Philip C. Cooley, and Donald S. Burke. 2006. "Strategies for Mitigating an Influenza Pandemic." *Nature* 442, no. 7101: 448–52.

Ferguson, Neil M., Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, and others. 2020. *Impact of Non-pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand*. London: Imperial College.

Ferretti, Luca, Chris Wymant, James Petrie, Daphne Tsallis, Michelle Kendall, Alice Ledda, and others. 2023. "Digital Measurement of SARS-CoV-2 Transmission Risk from 7 Million Contacts." *Nature* 626, no. 7997: 145–50.

Fraser, Christophe, Derek A. T. Cummings, Don Klinkenberg, Donald S. Burke, and Neil M. Ferguson. 2011. "Influenza Transmission in Households during the 1918 Pandemic." *American Journal of Epidemiology* 174, no. 5: 505–14.

Fraser, Christophe, Steven Riley, Roy M. Anderson, and Neil M. Ferguson. 2004. "Factors That Make an Infectious Disease Outbreak Controllable." *Proceedings of the National Academy of Sciences* 101, no. 16: 6146–51.

Funk, Sebastian, Shweta Bansal, Chris T. Bauch, Ken T. D. Eames, W. John Edmunds, Alison P. Galvani, and Petra Klepac. 2015. "Nine Challenges in Incorporating the Dynamics of Behaviour in Infectious Diseases Models." *Epidemics* 10: 21–25.

Gans, Joshua S. 2022. "The Economic Consequences of R = 1: Towards a Workable Behavioural Epidemiological Model of Pandemics." *Review of Economic Analysis* 14, no. 1: 3–25.

George, Dylan B., Wendy Taylor, Jeffrey Shaman, Caitlin Rivers, Brooke Paul, Tara O'Toole, and others. 2019. "Technology to Advance Infectious Disease Forecasting for Outbreak Management." *Nature Communications* 10, article no. 3932.

Germann, Timothy C., Kai Kadau, Ira M. Longini Jr., and Catherine A. Macken. 2006. "Mitigation Strategies for Pandemic Influenza in the United States." *Proceedings of the National Academy of Sciences* 103, no. 15: 5935–40.

Gimma, Amy, James D. Munday, Kerry L. M. Wong, Pietro Coletti, Kevin van Zandvoort, Kiesha Prem, and others. 2022. "Changes in Social Contacts in England during the COVID-19 Pandemic between March 2020 and March 2021 as Measured by the CoMix Survey: A Repeated Cross-Sectional Study." *PLoS Medicine* 19, no. 3: e1003907.

Hamner, Lea, Polly Dubbel, Ian Capron, Andy Ross, Amber Jordan, Jaxon Lee, and others. 2020. "High SARS-CoV-2 Attack Rate following Exposure at a Choir Practice—Skagit County, Washington, March 2020." *Morbidity and Mortality Weekly Report* 69, no. 19: 606–10.

Hatchett, Richard J., Carter E. Mecher, and Marc Lipsitch. 2007. "Public Health Interventions and Epidemic Intensity during the 1918 Influenza Pandemic." *Proceedings of the National Academy of Sciences* 104, no. 18: 7582–87.

Hellewell, Joel, Sam Abbott, Amy Gimma, Nikos I. Bosse, Christopher I. Jarvis, Timothy W. Russell, and others. 2020. "Feasibility of Controlling COVID-19 Outbreaks by Isolation of Cases and Contacts." *The Lancet Global Health* 8, no. 4: e488–96.

Hollingsworth, T. Déirdre, Don Klinkenberg, Hans Heesterbeek, and Roy M. Anderson. 2011. "Mitigation Strategies for Pandemic Influenza A: Balancing Conflicting Policy Objectives." *PLoS Computational Biology* 7, no. 2: e1001076.

Jia, Katherine M., William P. Hanage, Marc Lipsitch, Amelia G. Johnson, Avnika B. Amin, Akilah R. Ali, Heather M. Scobie, and David L. Swerdlow. 2023. "Estimated Preventable COVID-19-Associated Deaths due to Non-vaccination in the United States." *European Journal of Epidemiology* 38, no. 11: 1125–28.

Jones, Jefferson M., Mars Stone, Hasan Sulaeman, Rebecca V. Fink, Honey Dave, Matthew E. Levy, and others. 2021. "Estimated US Infection- and Vaccine-Induced SARS-CoV-2 Seroprevalence Based on Blood Donations, July 2020–May 2021." *JAMA* 326, no. 14: 1400–1409.

Jung, Sung-mok, Sara L. Loo, Emily Howerton, Lucie Contamin, Claire P. Smith, Erica C. Carcelén, and others. 2024. "Potential Impact of Annual Vaccination with Reformulated COVID-19 Vaccines: Lessons from the U.S. COVID-19 Scenario Modeling Hub." *PLoS Medicine* 21, no. 4: e1004387.

Kerpen, Phil, Stephen Moore, and Casey B. Mulligan. 2022. "A Final Report Card on the States' Response to COVID-19." *International Journal of the Economics of Business* 30, no. 2: 139–58.

Kissler, Stephen M., Christine Tedijanto, Edward Goldstein, Yonatan H. Grad, and Marc Lipsitch. 2020. "Projecting the Transmission Dynamics of SARS-CoV-2 through the Postpandemic Period." *Science* 368, no. 6493: 860–68.

Klaassen, Fayette, Melanie H. Chitwood, Ted Cohen, Virginia E. Pitzer, Marcus Russi, Nicole A. Swartwood, Joshua A. Salomon, and Nicolas A. Menzies. 2023. "Changes in Population Immunity against Infection and Severe Disease from Severe Acute Respiratory Syndrome Coronavirus 2 Omicron Variants in the United States between December 2021 and November 2022." *Clinical Infectious Diseases* 77, no. 3: 355–61.

Koutsakos, Marios, Adam K. Wheatley, Karen Laurie, Stephen J. Kent, and Steve Rockman. 2021. "Influenza Lineage Extinction during the COVID-19 Pandemic?" *Nature Reviews Microbiology* 19, no. 12: 741–42.

Lipsitch, Marc, Mary T. Bassett, John S. Brownstein, Paul Elliott, David Eyre, M. Kate Grabowski, and others. 2023. "Infectious Disease Surveillance Needs for the United States: Lessons from COVID-19." ArXiv:2311.13724. Ithaca, N.Y.: Cornell University.

Longini, Ira M., Jr., Azhar Nizam, Shufu Xu, Kumnuan Ungchusak, Wanna Hanshaoworakul, Derek A. T. Cummings, and M. Elizabeth Halloran. 2005. "Containing Pandemic Influenza at the Source." *Science* 309, no. 5737: 1083–87.

Meltzer, Martin I., Manoj Gambhir, Charisma Y. Atkins, and David L. Swerdlow. 2015. "Standardizing Scenarios to Assess the Need to Respond to an Influenza Pandemic." *Clinical Infectious Diseases* 60, no. suppl. 1: S1–S8.

Mills, Christina E., James M. Robins, and Marc Lipsitch. 2004. "Transmissibility of 1918 Pandemic Influenza." *Nature* 432, no. 7019: 904–6.

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

Ong, David S. Y., Paraskevi C. Fragkou, Valentijn A. Schweitzer, Roy F. Chemaly, Charalampos D. Moschopoulos, and Chrysanthi Skevaki. 2021. "How to Interpret and Use COVID-19 Serology and Immunology Tests." *Clinical Microbiology and Infection* 27, no. 7: 981–86.

Patrick, David M., Martin Petric, Danuta M. Skowronski, Roland Guasparini, Timothy F. Booth, Mel Krajden, and others. 2006. "An Outbreak of Human Coronavirus OC43 Infection and Serological Cross-Reactivity with SARS Coronavirus." *Canadian Journal of Infectious Diseases and Medical Microbiology* 17, no. 6: 330–36.

Pouwels, Koen B., Thomas House, Emma Pritchard, Julie V. Robotham, Paul J. Birrell, Andrew Gelman, and others. 2021. "Community Prevalence of SARS-CoV-2 in England from April to November 2020: Results from the ONS Coronavirus Infection Survey." *The Lancet Public Health* 6, no. 1: e30–e38.

Shaman, Jeffrey, and Marta Galanti. 2020. "Will SARS-CoV-2 Become Endemic?" *Science* 370, no. 6516: 527–29.

Stadlbauer, Daniel, Jessica Tan, Kaijun Jiang, Matthew M. Hernandez, Shelcie Fabre, Fatima Amanat, and others. 2021. "Repeated Cross-Sectional Sero-Monitoring of SARS-CoV-2 in New York City." *Nature* 590, no. 7844: 146–50.

Ueda, Peter, Catherine H. Mercer, Cyrus Ghaznavi, and Debby Herbenick. 2020. "Trends in Frequency of Sexual Activity and Number of Sexual Partners among Adults Aged 18 to 44 Years in the US, 2000–2018." *JAMA Network Open* 3, no. 6: e203833.

Velde, François R. 2022. "A Model of Economic Activity in San Francisco during the 1918 Influenza Epidemic." Working Paper 2022-04. Federal Reserve Bank of Chicago. https://www.chicagofed.org/publications/working-papers/2022/2022-04.

Wesolowski, Amy, Caroline O. Buckee, Kenth Engø-Monsen, and Charlotte Jessica Eland Metcalf. 2016. "Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data." *Journal of Infectious Diseases* 214, suppl. 4: S414–20.

Zhang, Xu-Sheng, Sema Mandal, Hamish Mohammed, Charlie Turner, Isaac Florence, Josephine Walker, and others. 2024. "Transmission Dynamics and Effect of Control Measures on the 2022 Outbreak of Mpox among Gay, Bisexual, and Other Men Who Have Sex with Men in England: A Mathematical Modelling Study." *The Lancet Infectious Diseases* 24, no. 1: 65–74.

Comments and Discussion

COMMENT BY

THOMAS PHILIPPON Atkeson and Kissler provide an important analysis of how vaccines and behavioral changes saved lives during the COVID-19 epidemic. They estimate that the vaccines and associated behavioral responses saved 800,000 lives compared to a counterfactual with no vaccine.

This is a large effect. The lives saved represent approximately 0.25 percent of the US population. As a comparison, the H1N1 epidemic of 1918–1919 killed about 0.65 percent of the US population.

A key result of the paper is that there is a strong complementarity between vaccines and behavioral responses. Without any change in behavior, the death toll would have been close to 1 percent of the population, in large part because many people would have become infected before vaccines became available. By contrast, Atkeson and Kissler estimate that roughly two-thirds of the population got vaccinated before their first infection and that vaccines were very efficient, lowering fatality by a factor of five at the peak of the epidemic in 2021.

THE ROLE OF VACCINES Atkeson and Kissler use granular data to document some key facts and develop a structural epidemiological model to interpret the facts and compute counterfactuals.

Their data come from three sources: serology data to keep track of immunity following infections and vaccinations; deaths associated with COVID-19 over time and across regions; and data linking vaccines and mortality from thirty US states.

The first key estimate is vaccines saved about 800,000 lives, as can be seen from the difference between the first and second lines of table 1 in the

Brookings Papers on Economic Activity, Spring 2024: 113–132

paper. It contains the estimates from the structural model, but a simple back-of-the-envelope calculation is also possible. The serology data suggest that essentially everyone has been infected at least once by now, but that 68 percent of the population received a vaccine before their first infection. The US population in 2020 was around 330 million, and the authors estimate that 1.2 million people have died from COVID-19.

The infection fatality rate (IFR) changes a lot over time for a variety of reasons: new variants, better treatments, vaccines, and so on. In 2020, the virus infected 11.5 percent of the population and killed 390,000 people, which implies an IFR around 1 percent. In 2021, the IFR decreased to 0.66 percent. After 2021, it was around 0.2 percent. Vaccines reduced the IFR by a factor of more than five during 2021 but much less afterward.

Atkeson and Kissler capture these changes in several ways. In the pre-Omicron period, they assume that the IFR falls over time, starting from a high value of 1 percent in early 2020 and eventually decreasing to 0.5 percent, as indicated by the serology data. When the first wave of Omicron arrives with its large increase in infections but a smaller increase in deaths, the implied IFR decreases further to 0.15 percent.[1]

In the authors' baseline calibration, the average IFR over the sample period is then 0.5 percent for the "naive unvaccinated" and 0.13 percent for vaccinated people. A rough estimate of lives saved is then the difference in IFR applied to the population that was vaccinated before the first infection: $(0.5\% - 0.13\%) \times 0.68 \times 330$ million = 830,000. This is in the ballpark of the more precise estimate from the structural model.

THE "SURPRISINGLY" LARGE IMPACT OF BEHAVIORAL CHANGES The second key takeaway from the serology survey data is that mitigation played a crucial role during 2020. By January 2021, less than 20 percent of the population had been infected. By contrast, model simulations predict that essentially the entire population would have been infected without a behavioral response. Together with the 1 percent IFR discussed earlier, this would have led to more than 3 million deaths.

The large role of behavioral changes provides a strong motivation for the development of the structural model. The structural model allows Atkeson and Kissler to study counterfactual experiments that would otherwise be unknowable.

The model is quite advanced and granular. It features waning immunity and takes into account the appearance of variants (Alpha, Delta, and

1. The 0.15 percent applies to susceptible agents (S). The authors also allow breakthrough infections from Omicron (in the R population) but with a very low IFR.

Omicron). A standard susceptible-infectious-recovered (SIR) model splits the population into three groups: susceptible, infectious, and resistant. The authors add two more groups: exposed but not yet infectious (E) and hospitalized (H). They also account for people who are vaccinated prior to their first infection. The E and I groups are indexed by the variant of the virus, and the R group is subject to (rare) breakthrough infections from the Omicron variant.

With this model Atkeson and Kissler obtain reliable estimates of the likely death rates under alternative scenarios. They then argue convincingly that the behavioral response was much stronger and longer lasting than in previous epidemics.

Should we then call it a surprise? I suppose it all depends on the relevant information set. A surprise is, by definition, the difference between an outcome and its expectation based on the prior information set. If we take as our information set the average strength and duration of behavioral responses in previous epidemics, as illustrated by the surveys published before the pandemic, then we must agree with the authors.

I would argue, however, that one should include both the existence of the internet and of the welfare state in our information set. Jones, Philippon, and Venkateswaran (2021) show that the possibility to work and shop remotely had a large impact on COVID-19 mitigation and saved approximately 200,000 lives. These options did not exist in the past, but they were (at least partly) predictable.

Similarly, the 1918 influenza occurred before the expansion of the welfare state. For many households, not working to slow down the spread of the virus would have meant extreme hardship. The $5 trillion fiscal response (Romer 2021) would have been simply unimaginable at the time. The social insurance and public health components of the fiscal response to COVID-19 afforded households the possibility to reduce in-person labor supply, even though the rest of the spending was arguably superfluous (Romer 2021).

PREPARING AGAINST FUTURE CRISES Atkeson and Kissler argue that, to prepare for the next epidemic, we must improve data collection and analytics. We must determine the pathogen's transmission routes and keep track of incidence and immunity.

I fully agree with these points, but I would add several nonpharmaceutical interventions to the list. A striking feature of the COVID-19 epidemic is its unequal impact across groups and locations. As is well known, the virus was ten times more dangerous for old people than for young people. Similarly, Jones, Philippon, and Venkateswaran (2021) show that the risk of infection varied by a factor of five across occupations. Finally, we see from the current paper that outcomes also differed by a factor of five across states: in terms of fatality rates, New Hampshire and Vermont look like Denmark, while Arizona and Mississippi look like Russia.

These large differences across demographic groups, occupations, and locations imply that we can reduce the severity of future pandemics with targeted interventions. An obvious one is to improve options for remote work and schooling, starting with universally available broadband internet. While some tasks cannot be done remotely, the large differences in exposures across occupations suggest that significant improvements are possible. Similarly, regarding schooling, while we know that in-person teaching is preferable, it seems likely that some form of remote learning will be needed in future crises, and it is therefore important to ensure equal access to computers and reliable internet connections for all students.

Differences across locations are harder to interpret since they reflect differences in preferences as well as governance choices for given preferences. The scale of differences in fatality rates, however, suggests that differences in preferences are unlikely to account for all the variation that we observe across states. Learning and emulating best practices can therefore improve the policy trade-off between mitigation, individual freedoms, and economic damages.

REFERENCES FOR THE PHILIPPON COMMENT

Jones, Callum, Thomas Philippon, and Venky Venkateswaran. 2021. "Optimal Mitigation Policies in a Pandemic: Social Distancing and Working from Home." *Review of Financial Studies* 34, no. 11: 5188–223.

Romer, Christina D. 2021. "The Fiscal Policy Response to the Pandemic." *Brookings Papers on Economic Activity*, Spring: 89–110.

COMMENT BY

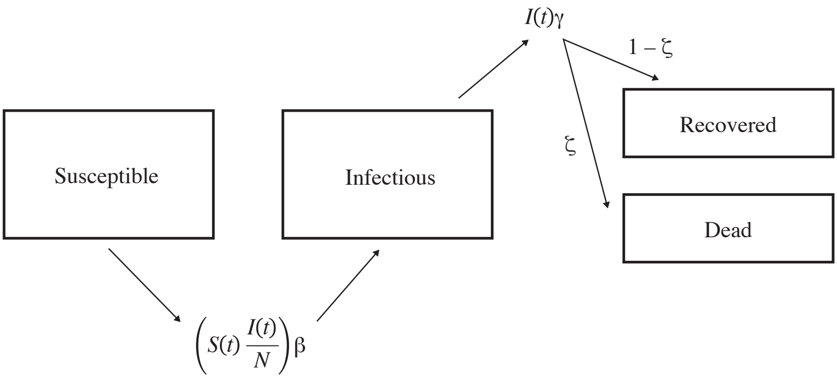
COADY WING The COVID-19 pandemic upended life in the United States. People changed their behaviors to mitigate the risk of infection and mortality. Governments imposed new regulations designed to encourage further reductions in the transmission of the virus and to promote vaccine take-up once the vaccine became available. By now, there is a large body of literature in social and health sciences that documents these behavioral responses and tries to evaluate the intended and unintended consequences of various policy initiatives (Gupta and others 2021b; Autor and others 2022; Chetty, Friedman, and Stepner 2024).

In their paper, Atkeson and Kissler evaluate the high-level effect of the pandemic response on the level and time series dynamics of COVID-19 mortality. They use a structural epidemiological model to decompose the way that mortality is determined by transmission-related behaviors, vaccine take-up, and shifts in transmissibility and virulence of the virus. This is a compartmental model in which a population of susceptible people transitions through a collection of health states. Transitions are governed by model parameters that define—at each point in time—the viral transmission rate, the duration of infectiousness, the take-up and effectiveness of the vaccine, and the infection fatality rate. The parameters of the model are not estimated from the data. In some cases, the authors draw on epidemiological studies to guide the choice of parameters that represent infection fatality rates of different strains. But for the most part, Atkeson and Kissler choose parameterizations that seem plausible, most likely using some amount of trial and error. The main evidence that the chosen parameters are sensible is that the model does an excellent job of reproducing the observed time series of COVID-19 mortality in the United States. It also fits the time series estimates of the cumulative share of the population that had been infected by COVID-19 based on convenience samples from the Blood Donor Survey and Commercial Lab Survey.

Treating the model as correct, Atkeson and Kissler examine counterfactual scenarios to measure the role of specific mechanisms in causing mortality. For example, in one scenario, behavioral changes are maintained but the vaccine never arrives. In another simulation, people do not engage in major behavioral changes, but the vaccine becomes available on schedule. In both cases, they use the model to compute the number of COVID-19 deaths that would have occurred if the parameters of the model are correct but certain events played out differently. The simulations suggest that the combination of behavioral changes and the eventual availability of the vaccine led to substantial reductions in mortality. Without behavioral changes, the vaccine would have arrived too late to matter. Without the vaccine, the behavioral changes would mostly have reallocated deaths over time. One way to see it is that behavioral changes that mitigate transmission early in the pandemic increase the marginal health benefits of vaccines later in the pandemic. Behavior and vaccines are complements in an aggregate health production function.

In my discussion, I focus on three main topics. First, I try to provide an intuitive account of the type of model that Atkeson and Kissler use in their analysis and to point out some of the key assumptions involved in such models. Second, I discuss some of the ways we might judge the credibility

Figure 1. The Classic SIR Model



Source: Author's illustration.

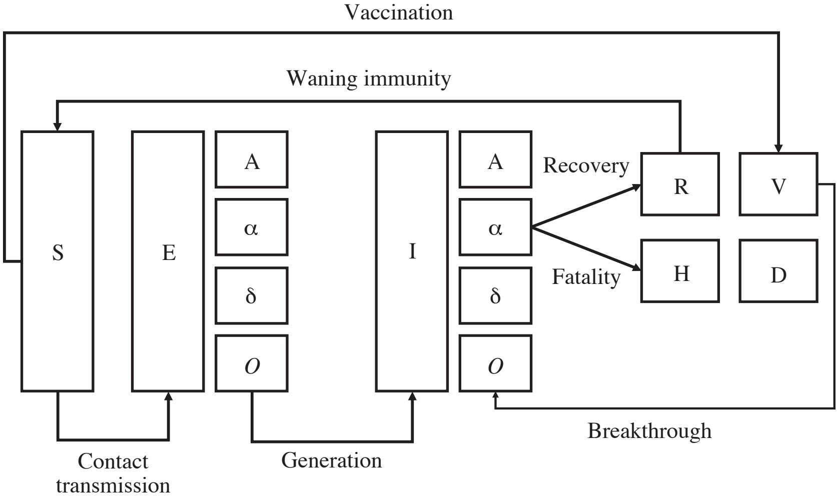
of structural epidemiological models and suggest some ideas for incorporating quasi-experimental study designs. Third, I shift focus to questions about the behavioral determinants of vaccine take-up with particular attention to take-up coming out of the COVID-19 pandemic. I present some early research examining the way that vaccine take-up may be undermined by breakthrough infections.

STRUCTURAL EPIDEMIOLOGICAL MODELS The most famous model of an epidemic is the susceptible-infectious-recovered (SIR) model, which was developed by Kermack and McKendrick (1927). Figure 1 is a visual representation of a basic SIR model. In this setup, the whole population starts out susceptible to the disease except for a single index patient who is infected. The people in the susceptible and infectious compartments mix at random. When an infectious person and a susceptible person come into contact, the susceptible person is infected at rate β. Infected people recover and leave the infectious compartment according to the recovery rate, γ. The infection mortality rate is ζ.

The share of the population in each compartment at a given point in time is determined by these transmission, recovery, and mortality parameters. In a setting where the whole population is susceptible, an infected person generates $R = \frac{\beta}{\gamma}$, new infections. When $R < 1$, the disease dies out over time. When $R > 1$, there is an exponential outbreak in the number of new infections.

Real-world conditions are more complicated than the simple model implies. For example, the population of susceptible and infectious people

Figure 2. A Stylized Version of Atkeson and Kissler's SEIHRS Model



Source: Author's illustration.

might be structured so that some people have higher mixing rates than others, violating the random mixing assumption of the basic model. The basic SIR model is not capable of representing epidemics that exhibit repeated waves of infections and in which the properties of the pathogen itself might change over time. And—of course—economists often point out that the simple SIR framework does not allow for people who change their behaviors in response to prevailing epidemiological conditions, new public policies, or expectations about future technologies such as vaccines or cures.

The contemporary literature has elaborated and complexified the basic SIR model to make it more realistic. A visual representation of Atkeson and Kissler's model would look something like figure 2. The first thing to notice is that there are many more compartments. At a high level, the model is organized around six main states: susceptible, exposed, infectious, hospitalized, resistant, and dead. But the exposed and infectious compartments are subdivided further by the viral strain—there is a box each for the original ancestral strain as well as the Alpha, Delta, and Omicron viral variants. Each of the strains has a different transmission rate and infection mortality rate. The resistant compartment includes space for people who have recovered from an actual infection and for people who have been vaccinated. The flow of people through the compartments is no longer in a single direction. Immunity from infection may wane over time, creating a flow of people from the resistant to the susceptible. Vaccinated people may experience breakthrough infections.

Simulating the model involves specifying the timing of certain shocks to the system. Some pathways open up at particular calendar times because a new viral strain emerges. The parameters governing transmission and mortality change over time too. Taken together, the model simulations involve a schedule of epidemiological shocks (new viral variants), technology shocks (vaccine availability), seasonality, behavioral relationships (transmission-mortality elasticity), and structural changes in behavioral relationships (fatigue).

To get a feel for how the model works, it helps to consider the block of the model that shapes the transmission rate at a point in time. In Atkeson and Kissler's model, the transmission rate at a point in time follows a seasonal component and a behavioral component in which the transmission rate responds to the number of daily deaths. The idea is that when death rates get high, people change behavior to reduce infection risk. After a time, fatigue sets in and the responsiveness of transmission to fatalities shrinks. Formally, the transmission relationship in the model is:

$$\beta_j(t) = \bar{\beta}_j \times \exp\left[-\kappa(t)\frac{dD(t)}{dt} + \psi(t)\right].$$

In this expression, $\bar{\beta}_j$ is the inherent transmissibility of viral strain j, $\psi(t)$ is the effect of the season on transmission, $\frac{dD(t)}{dt}$ is the level of daily deaths, and $\kappa(t)$ is the semi-elasticity of transmission with respect to deaths. In practice, the seasonal function $\psi(t)$ is a cosine wave parameterized to match fall-winter versus spring-summer patterns. For the original COVID-19 strain, the authors set $\bar{\beta}_j = 1.2$. To represent the behavioral component of transmission, they set the baseline semi-elasticity to be $\bar{\kappa} = 250{,}000$. To incorporate the idea of behavioral fatigue, they smoothly shrink the behavioral response down to $.35 \times \bar{\kappa}$ by late November 2020. Once a set of parameters has been chosen, Atkeson and Kissler run their model, pushing an initial population through the various compartments and keeping track of how many people are dead, infectious, and vaccinated at each point in time.

IDENTIFICATION PROBLEMS AND EPIDEMIOLOGICAL MODELS Constructing epidemiological models is challenging for some of the same reasons that constructing macroeconomic models is challenging. The parameters of the model are hard to cleanly identify because the variation generated during real-world outbreaks is not randomized across places, time periods, and people. And the outcomes realized during an epidemic seem very contingent on a high-dimensional set of conditions and constraints. This makes it hard to accumulate knowledge across different settings or to discriminate between one hypothesized model and another.

How should we judge the credibility of Atkeson and Kissler's model? One natural strategy is to compare the outputs of the model with observed outcomes from the real world. That is the approach that Atkeson and Kissler take. Figures 2 and 4 from their paper show a tight correspondence between COVID-19 mortality and infections as generated by the model and the actual time series data on COVID-19 mortality and COVID-19 infections. Comparing the model-predicted mortality and infection series with their real-world counterparts is essentially the same idea macroeconomists use when they form judgments based on how well a specific model is able to match moments observed in the real world.

Although the close fit shown in Atkeson and Kissler's figures 2 and 4 is impressive, a fundamental question is whether this collection of compartments and parameters is really a good representation of the process that generated those deaths and infections. Are there other combinations of parameters and compartments that might also fit the data very well but could generate quite different counterfactual simulations? To what extent is the good fit of the model akin to a regression that fits the data well in sample but performs badly at out-of-sample forecasts?

The model makes strong assumptions about the mechanism and even the specific numerical values of key causal parameters. The payoff from these kinds of assumptions is substantial: you can use the model to simulate the pandemic under alternative conditions, which is just the type of thing you need to do to study alternative policy options. But the credibility of the counterfactual simulations depends on the plausibility of the underlying modeling choices. For example, is $\bar{\kappa} = 250{,}000$ a plausible value for the behavioral response to mortality in the early pandemic? Does this choice undershoot the degree to which transmission responded to mortality? It seems hard to decide something like this through intuition.

One idea is to combine the structural epidemiological methods with research strategies that are common in empirical microeconomics, which focus on identifying causal effects using plausibly exogenous variation. Nakamura and Steinsson (2018) discuss identification problems in macroeconomics, pointing out that macroeconomists often judge the credibility of specific models by their ability to match moments observed in the real world. Often the moments used in this type of work are simple aggregate means and variances. But Nakamura and Steinsson (2018) highlight that more recent research involves efforts to match *identified* moments. Identified moments are causal effects identified using the methods popular in empirical microeconomics: regression discontinuity designs, difference-in-differences designs, or instrumental variable designs. Taking advantage of identified moments to judge the performance of a more complex structural model or to pin down the value of a class of parameters from a structural model is a strategy that may be useful for future work on epidemiological models.

To take one very small step in this direction, we could compare estimates from Atkeson and Kissler's model with identified moments from related quasi-experimental studies. For example, Gupta and others (2021a) use a generalized difference-in-differences regression to make reduced-form estimates of the effects of the early vaccination campaign on cumulative COVID-19 mortality over the first five months of the vaccination campaign. Their estimates come from quasi-likelihood Poisson regression models with the following basic form:

$$M_{st} = \exp\left[\sum_{k=0\ldots4} \delta_k V_{st-k} + a_s + b_t\right] + e_{st}.$$

In the regression, M_{st} is the cumulative number of COVID-19 deaths per 100 adults in state s as of week t, and V_{st} is the cumulative number of doses administered per 100 adults in state s by week t. The model includes state and week fixed effects and is intended to measure the effects of the vaccine rollout by exploiting variation in the speed of vaccine distribution across states. The estimated parameters from this two-way fixed effects specification are used to estimate the counterfactual cumulative COVID-19 mortality rate in the absence of the vaccination campaign. The results imply that by the second week of May in 2021 the vaccination campaign had already averted about 139,393 COVID-19 deaths. How do these reduced-form estimates line up with the simulations from Atkeson and Kissler's model? The numbers underlying the left panel of their figure 5 imply that by the second week of May the vaccination campaign had averted 126,664 COVID-19 deaths. This is well inside the confidence interval of the estimate by Gupta and others (2021a), perhaps suggesting that Atkeson and Kissler's model fares pretty well at matching an identified moment that is a bit more removed from the analysis than the mortality time series itself.[1] Atkeson and Kissler's model-based estimates extend beyond the first few months of the campaign, and they suggest that the impact of the vaccination campaign continued to grow over the course of 2021 and then diminished in 2022. The model-based estimates imply that most of the additional deaths that would have occurred in the absence of the vaccine would have happened by the end of 2021. This is because without the vaccine nearly everyone would have been infected during the Omicron wave in early 2022.

Atkeson and Kissler are surely correct that people respond to epidemiological conditions and the availability of vaccines. And public policies designed to control an epidemic are premised on the idea that behavior is both malleable and an important determinant of the path of the epidemic. Their model provides an excellent example of how to integrate these policy relevant relationships into an epidemiological model. But these models would be more compelling if there were more quasi-experimental studies trying to pin down the specific ways that people respond to changing conditions and how those changes affect downstream population health outcomes. In particular, economists could be useful by developing the identification strategies and data sources needed to estimate things like: (a) the causal effect of mortality on disease transmission (behavioral responses); (b) the causal determinants of vaccine take-up and behavioral fatigue; and (c) the role of differentiated contact patterns on disease outcomes.

VACCINE TAKE-UP One of the main lessons that Atkeson and Kissler draw from their analysis is that behavioral adaptations that reduced transmission rates during the first year of the pandemic allowed the COVID-19 vaccine to substantially reduce overall mortality from COVID-19. Specifically, in simulations where they keep the behavioral parameters fixed but never turn on the availability and take-up of the vaccine, there would have been almost 800,000 additional COVID-19 deaths.

The implications of the model are quite encouraging in certain ways. They suggest that it is possible to use behavioral modifications to suppress a pandemic long enough to develop and distribute a vaccine soon enough for the vaccine to actually save lives. At the same time, as Atkeson and Kissler are careful to point out, there is a lot of historical contingency involved in this analysis. If the highly transmissible Omicron variant had

1. Not that removed, of course: the two-way fixed effects regression in Gupta and others (2021a) is based on state *x* week-level mortality data. Atkeson and Kissler are working with a national mortality time series rather than a state *x* week panel.

arrived sooner, then the vaccine likely would not have saved many lives: by the time the vaccine arrived, it would have been too late. On the other hand, if the variant had not appeared or had appeared even later, then the vaccine would have had even more of an impact. Governments have little influence on the appearance and characteristics of new viral strains and so the strategy of behavior-induced transmission reduction followed by vaccination is somewhat inherently risky. However, the distribution and take-up of the vaccine itself is something that may deserve more attention.

In particular, it would make sense for economists to develop a better understanding of the determinants of vaccine take-up both during an epidemic and during regular conditions. Neoclassical economics suggests that vaccine take-up may be *too low* from a social welfare point of view because vaccines may produce positive externalities. Acton and others (2022) studied college vaccine mandates and found some evidence that mandates led to lower rates of COVID-19 spread in nearby communities. Freedman and others (2023) use linked micro data on COVID-19 tests, vaccinations, and health care records to study vaccine spillovers in middle schools and households with children in Indiana. They find little evidence of spillovers in schools, but they do find vaccine spillovers in households. Since households might plausibly internalize these vaccine spillovers, it is not obvious that free-riding on positive externalities is a major determinant of low vaccine take-up.

Another explanation for low vaccine take-up is that people's assessment of the private net benefits of the vaccine is somewhat lower than expected. Recent work by Carlin and others (2022) used discrete choice survey experiments to measure people's willingness to pay to be vaccinated for COVID-19 during early 2021. They found that median willingness to pay was around \$50. Back-of-the-envelope calculations based on estimates of the value of statistical life from other contexts imply people should have been willing to pay around \$2,700 to be vaccinated, given the mortality effects of the vaccine and prevailing caseloads. Thus, people seem to undervalue the COVID-19 vaccine. This could be one explanation for relatively low vaccine take-up in the United States. People might undervalue vaccines for many different reasons, including concerns about the safety of the vaccine or the perceived costs of vaccine side effects, needle aversion, or the political symbolism of the vaccine.

Another possibility is that people's demand for a vaccine is partly derived from their own experience of the vaccine and the underlying illness. For example, Jin and Koch (2021) study the relationship between influenza vaccination and influenza infection at the individual level over time. They find that contracting influenza in one year affects people's take-up of the vaccine in future years, suggesting that people learn from suffering. However, they also find that what people learn depends on their vaccination history. People who were unvaccinated and infected in the baseline year are more likely to be vaccinated in the future. But this learning-induced demand for the vaccine is offset for people who were vaccinated and experienced a breakthrough infection. One interpretation is that breakthrough infections undermine people's assessment of the usefulness of a preventive vaccine. This is almost certainly the wrong conclusion for people to draw: breakthrough infections can and do occur even when the vaccine is effective. Nevertheless, this type of misguided behavioral response is not difficult to understand.

LEARNING BY SUFFERING IN INDIANA A similar dynamic may hold for the COVID-19 vaccine as people make choices about boosters and vaccine take-up in nonepidemic conditions. To shed some light on the issue, I used linked administrative data from Indiana to study the relationship between vaccine take-up and prior vaccination and COVID-19 infection experiences. There are three main data sources: (1) Indiana COVID-19 vaccination registry; (2) Indiana COVID-19 lab test registry; and (3) Indiana Network for Patient Care (INPC) research database. INPC is a database of electronic medical records contributed by most of the hospitals and clinics in Indiana. I constructed a study sample of people who had at least one health care encounter in the INPC system between 2018 and 2019. Then I linked these individual records with COVID-19 vaccination records and COVID-19 lab tests and test results from 2020 to 2022.

With the data in hand, I estimate simple cross-sectional regressions with the following form:

$$Vax_i^{2022} = \alpha_0 + \alpha_1 Vax_i^{2021} + \alpha_2 Covid_i^{2021} + \alpha_3 \left(Vax_i^{2021} \times Covid_i^{2021} \right) + X_i \beta + e_i .$$

In the regression, Vax_i^{2022} is a binary variable indicating whether person i received the COVID-19 vaccine in 2022. Vax_i^{2021} indicates whether the person was vaccinated in 2021, and $Covid_i^{2021}$ indicates whether the person had a lab-confirmed COVID-19 infection in 2021. X_i is a covariate vector that adjusts for gender, race, and age fixed effects. I fit an overall regression to the full sample as well as separate regressions for younger, middle aged, and older people. The estimated regression coefficeints in table 1 show that the 2022 vaccination rate is much lower than the vaccination rate during the main pandemic. The take-up rate in 2022 was about 12 percent among

Table 1. Regressions of COVID-19 Vaccine Take-Up on Prior Season COVID-19 Infection and Vaccination

	Age 18–39	*Age 40–64*	*Age 65+*
Intercept	–0.011***	0.008***	0.033***
	(0.002)	(0.001)	(0.001)
2021 infection	0.015***	0.023***	0.024***
	(0.001)	(0.001)	(0.001)
2021 vaccine	0.167***	0.119***	0.089***
	(0.001)	(0.001)	(0.001)
2021 infection × 2021 vaccine	–0.017***	–0.012***	–0.013***
	(0.002)	(0.002)	(0.002)
N	820,124	1,107,246	876,854
R^2	0.06	0.04	0.03
Mean of outcome	0.061	0.091	0.118

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Source: Author's calculations.

Note: The regressions adjust for gender-, race-, and age-fixed effects. Standard errors are estimated using a heteroskedasticity robust variance matrix.

people over age 65, 9 percent among middle-aged adults age 40–64, and 6 percent among younger adults age 18–39.

Vaccination in 2022 was much higher among people who were vaccinated in 2021, suggesting preferences for the vaccine in the past are a strong predictor of vaccination in subsequent seasons. The coefficient on the prior COVID-19 infection indicator is also positive and quite large. Among those age 40–64 and age 65 and older who did not get vaccinated in 2021, the 2022 vaccination rate was about 2.4 percentage points higher among people who contracted COVID-19 in 2021 than among people who did not contract COVID-19. This learning by suffering effect is a bit smaller—only 1.5 percentage points—among younger adults age 18–39. However, the coefficient on the interaction term is negative, suggesting that experiencing a breakthrough infection offsets the learning by suffering effect perhaps because it undermines confidence in the vaccine. For middle-aged and older adults, the breakthrough effect offsets the learning by suffering effect by about 50 percent. For younger adults the breakthrough effect offsets the learning by suffering effect by over 100 percent, completely undoing any induced demand from prior infection.

This analysis suggests that the dynamics of individual COVID-19 vaccination exhibit some of the same patterns reported by Jin and Koch (2021) for influenza. In particular, take-up of the vaccine is partly determined by firsthand experience of the disease, and breakthrough infections seem to reduce subsequent demand for the vaccine. This type of response to health shocks seems undesirable from a public health point of view. It would probably be better if people did not lower their opinion of the efficacy of the vaccine on the basis of their own recent health experiences. But it is not at all hard to understand how a breakthrough infection might be a salient event that does motivate behavioral changes. Understanding how people interpret and change their behavior in response to salient health events in their own lives or in the lives of other people in their family may be an important way to develop more realistic models of epidemics.

REFERENCES FOR THE WING COMMENT

Acton, Riley K., Wenjia Cao, Emily E. Cook, Scott A. Imberman, and Michael F. Lovenheim. 2022. "The Effect of Vaccine Mandates on Disease Spread: Evidence from College COVID-19 Mandates." Working Paper 30303. Cambridge, Mass.: National Bureau of Economic Research.

Autor, David, David Cho, Leland D. Crane, Mita Goldar, Byron Lutz, Joshua Montes, William B. Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz. 2022. "The $800 Billion Paycheck Protection Program: Where Did the Money Go and Why Did It Go There?" *Journal of Economic Perspectives* 36, no. 2: 55–80.

Carlin, Patrick, Brian E. Dixon, Kosali I. Simon, Ryan Sullivan, and Coady Wing. 2022. "How Undervalued Is the Covid-19 Vaccine? Evidence from Discrete Choice Experiments and VSL Benchmarks." Working Paper 30118. Cambridge, Mass.: National Bureau of Economic Research.

Chetty, Raj, John N. Friedman, and Michael Stepner. 2024. "The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data." *Quarterly Journal of Economics* 139, no. 2: 829–89.

Freedman, Seth M., Daniel W. Sacks, Kosali I. Simon, and Coady Wing. 2023. "Direct and Indirect Effects of Vaccines: Evidence from COVID-19." Working Paper 30550. Cambridge, Mass.: National Bureau of Economic Research.

Gupta, Sumedha, Jonathan Cantor, Kosali I. Simon, Ana I. Bento, Coady Wing, and Christopher M. Whaley. 2021a. "Vaccinations against COVID-19 May Have Averted up to 140,000 Deaths in the United States." *Health Affairs* 40, no. 9: 1465–72.

Gupta, Sumedha, Thuy D. Nguyen, Shyam Raman, Byungkyu Lee, Felipe Lozano-Rojas, Ana I. Bento, Kosali I. Simon, and Coady Wing. 2021b. "Tracking Public and Private Responses to the COVID-19 Epidemic." *American Journal of Health Economics* 7, no. 4: 361–404.

Jin, Ginger Zhe, and Thomas G. Koch. 2021. "Learning by Suffering? Patterns in Flu Vaccination Take-Up." *American Journal of Health Economics* 7, no. 1: 68–94.

Kermack, William Ogilvy, and A. G. McKendrick. 1927. "A Contribution to the Mathematical Theory of Epidemics." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 115, no. 772: 700–721.

Nakamura, Emi, and Jón Steinsson. 2018. "Identification in Macroeconomics." *Journal of Economic Perspectives* 32, no. 3: 59–86.

GENERAL DISCUSSION James Stock noted an important contribution of the paper is a fairly simple model given the complexity of the task. On the interaction of behavior and vaccination, Stock reflected on the importance of waiting: first, and rather obvious, it gives you time to get the vaccine; second, given the high fatality rate in the first wave, those who took self-protective measures and waited and then became infected during later waves were better off. Stock further pointed to the notable result in figure 10 of the paper's conference draft, which shows the convergence of the effective reproduction number to one for every state, despite significant state differences in political views and COVID-19-related interventions.[1] This, he argued, points to the extensive self-protection measures taken by individuals across all states. Finally, Stock underscored the importance of continued work in this area, including better data collection, to better prepare us for similar events in the future.

Speaking to the cross-state variation in COVID-19 incidence, Louise Sheiner agreed with the authors' emphasis on the importance of behavior in explaining these differences. She argued that politicization was a major contributor, pointing to data that show that the variation in the labor force participation rate, unemployment, and consumption can be explained by political affiliation—the share of the state population that voted for Joe Biden. The same goes for state variation in vaccination rates. Sheiner concluded that this underscores people's attitudes as the primary driver of differences, rather than state lockdowns and other mandates during this time.

On the behavioral response, Carol Graham suggested that the authors further explore the great variation within states in vaccination rates—that observed between counties. Looking at the standard deviation within states, for example, would be useful in trying to better understand the outcome that seems to show there is convergence across states in the aggregate.

Stefanie Stantcheva asked the authors whether they had explicitly taken the multidose nature of the COVID-19 vaccine into account in their model, and if not, how such dynamics may alter the findings. Stantcheva then raised the role of trust in government as a salient issue to consider when contemplating future responses to a pandemic.

Tying together issues of data collection and trust in public officials, Steven Davis argued that there is a need to make data both transparent and credible to avoid politicization. Davis then turned to discuss the economic

1. This figure is included as figure B.6 in the online appendix of the paper's final version published in this *BPEA* volume.

resilience value of working from home. In his paper with Barrero and Bloom, survey results where respondents were asked if their internet connectivity affected their ability to work productively when working from home showed that while 75 percent said their internet connection was perfectly adequate, the other 25 percent reported that their productivity was negatively affected by the poor quality of their internet connection.[2] The paper estimates that subpar internet connectivity lowered US labor productivity by 3 percent in the period from May 2020 to April 2021. Davis also suggested that people may also be more willing to undertake voluntary efforts to slow the transmission of a virus if they can productively work from home, further underscoring the increased economic resilience that widespread access to high-quality internet would provide.

Kenneth Rogoff wondered how the model from the paper could be used to learn more about the economic cost of different choices. He observed that the learning losses for students during COVID-19 were stunning. What kind of information can we gather to help us better, and more quickly, determine what the right mitigation efforts are while we wait for a vaccine the next time around?

Maurice Obstfeld appreciated the contribution of the authors' work in successfully pinning down the role that delay played in the spread of COVID-19, particularly in the current politicized environment. However, Obstfeld worried that the next pandemic could be very different and called for better preparedness in general. We were lucky this time, Obstfeld reasoned, that mRNA technology was available, speeding up the rollout of vaccinations notably. He added that future pandemics are likely to be just as politicized, raising issues such as what the optimal approach to school closures would be if younger demographics were more vulnerable, as was the case with the 1918 influenza epidemic. Obstfeld brought up previous efforts, including two panels in 2021 (the Independent Panel for Pandemic Preparedness and Response and the High-Level Independent Panel) focused on funding and global cooperation on pandemic surveillance and response, as well as a book by Bill Gates, *How to Prevent the Next Pandemic*, on the topic. But he noted that few of the recommendations had been implemented since, and he cautioned that international attention to this issue had dwindled significantly with access to antivirals and the slowing fatality of

2. Jose Maria Barrero, Nicholas Bloom, and Steven J. Davis, "Internet Access and Its Implications for Productivity, Inequality, and Resilience," In *Rebuilding the Post-Pandemic Economy*, edited by Melissa S. Kearney and Amy Ganz (Washington: Aspen Institute Press, 2021). https://www.economicstrategygroup.org/publication/barrero-bloom-davis/.

COVID-19. Obstfeld stressed the need for funding and international cooperation to continue the work, including surveillance of animal reservoirs and thinking about what types of vaccines may be necessary depending on specific viruses of future pandemics.

Andrew Atkeson recalled how economists had sprung into action at the onset of the pandemic, thinking about sectoral-level interventions and testing, among other things. He firmly believed that it was not for lack of ideas that things did not happen. He pointed out the sometimes lackluster response of the public health sector and epidemiologists to the ideas of economists during this time. Atkeson stated that, just like the military, we need to plan for all types of contingencies and be ready with a response no matter the circumstances, noting that the willingness to spend given the economic cost in these situations is very high.

To Obstfeld's point, Şebnem Kalemli-Özcan agreed that the next pandemic could be very different but said there are still lessons to be drawn from COVID-19. She called for better financial targeting of funds in general. Kalemli-Özcan acknowledged that while mandated lockdowns may be a second-best option, relying on behavioral responses to mitigate the spread of a virus may not be feasible—especially in countries where a greater share of the labor force is informal. She suggested that to make lockdowns more efficient and financially sustainable, funds should be targeted to specific sectors: in the case of COVID-19, contact-intensive sectors. She further highlighted some of her own work on the topic and agreed that more data collection was needed.[3]

Stan Veuger cautioned that some of the policy suggestions in the paper might be difficult to implement. Mandating testing, he believed, would likely face opposition in the United States as well as in Western Europe. On data collection, Veuger was skeptical of the role the Centers for Disease Control and Prevention (CDC) would be able to play. He pointed to the absence of a US-wide representative survey of COVID-19 incidence despite a significant amount of additional funding to CDC during the pandemic.[4]

3. Cem Çakmaklı, Selva Demiralp, Şebnem Kalemli-Özcan, Sevcan Yeşiltaş, and Muhammed A. Yıldırım, "The Economic Case for Global Vaccinations: An Epidemiological Model with International Production Networks," working paper 28395 (Cambridge, Mass.: National Bureau of Economic Research, 2024).

4. Congressional Research Service, *US Public Health Service: COVID-19 Supplemental Appropriations in the 116th Congress* (Washington: Author, 2021), https://crsreports.congress.gov/product/pdf/R/R46711/3; and *American Rescue Plan Act of 2021 (P.L. 117-2): Public Health, Medical Supply Chain, Health Services, and Related Provisions* (Washington: Author, 2021), https://crsreports.congress.gov/product/pdf/R/R46834.

Atkeson said that they had purposefully tried to stay clear of the politics but agreed that we may not want federal mandates. He argued that we probably ought to accept the different decisions of democratically elected state officials, some of whom, during the COVID-19 pandemic, chose to accept higher death rates in their states.

Laurence Ball asked how events might have transpired if things had been a lot better than they were. How many fewer deaths could there have been had we pursued the optimal policy? Atkeson responded by suggesting that in the absence of vaccines being made available earlier, there was probably not much we could have done better.

Tristan Reed asked about the external validity of the authors' findings for developing countries, saying that most of the delay in vaccine delivery to developing countries during the pandemic could be attributed to them ordering later.[5] This may seem highly irrational at first, Reed conceded, and justifying not buying a vaccine seems to suggest a very low statistical value of life. However, Reed explained, the authors point to results in the paper that could suggest that in the absence of work-from-home technology, purchasing a vaccine is not worth it anymore after 120 days. He pondered whether when developing countries today are asked to financially contribute toward being able to purchase vaccines on day zero of a future pandemic and interest is muted, it reflects their belief that they do not have work-from-home technology.

On the low uptake of vaccines in emerging markets and its interaction with the need for better data, Raghuram Rajan talked about the specific case of India. He explained that India severely undercounted the deaths resulting from COVID-19. Initially, there were false stories spread about natural immunity against COVID-19 in India. Rajan said that the actual number of fatalities was suppressed, as the initial miscalculations would have negatively reflected on the capability of the public health system in each state. The second COVID-19 wave hit India hard because of the lack of immunization. In the official statistics, death rates in India are low, but taking undercounting into account raises the death toll significantly. In conclusion, Rajan said that this highlights the dangers of working with poor data and the policies made based on such data.

Hoyt Bleakley expressed his preference for adding standard tools of public economics: weighing marginal cost versus marginal benefit, including

5. Ruchir Agarwal and Tristan Reed, "Financing Vaccine Equity: Funding for Day-Zero of the Next Pandemic," *Oxford Review of Economic Policy* 38, no. 4 (2022): 833–50.

external cost and benefits. He agreed with the discussants' point that the benefits of getting vaccinated are largely internal, even though measuring the external benefits would be key for the design of a subsidy. Bleakley suggested using the model to measure not just the externality on the next person who could get infected, but also on the social and private incentives to delay infection before a vaccine becomes available. He also suggested looking at a much faster transmission rate in the authors' model, which would resemble that of Omicron at the onset of the pandemic, to see how that would alter the effect described in the paper.

Gerald Cohen noted that the discussion on the current paper and the paper by Stantcheva on inflation (also included in this *BPEA* volume) both focused on information available to the public: the current paper on the extent to which the public understood the propagation mechanisms for disease transmission during COVID-19 and Stantcheva's paper on whether the public are able to think correctly about inflation in general. Cohen suggested that economists should think about the importance of how to optimize people's information about the benefits of vaccines or private efforts of mitigation versus the benefits of the public better understanding inflation.

Speaking to the monetary losses during COVID-19, Robert Hall observed that a great number of workers were on temporary layoff around April 2020, and that output losses were large.[6] This came with a partially offsetting increase in leisure. He noted that there was no material decline in consumption. How do we put prices on these developments? Hall proposed that we carefully consider these different pieces to the puzzle in trying to measure the net loss of well-being from COVID-19.

6. According to Bureau of Labor Statistics (BLS), the number of unemployed people on temporary layoff reached 18 million in April 2020. BLS, "Temporary Layoffs Remain High following Unprecedented Surge in Early 2020," February 10, 2021, https://www.bls.gov/opub/ted/2021/temporary-layoffs-remain-high-following-unprecedented-surge-in-early-2020.htm.

SERKAN ARSLANALP
International Monetary Fund

BARRY EICHENGREEN
University of California, Berkeley

PETER BLAIR HENRY
Stanford University Hoover Institution

Sustained Debt Reduction: The Jamaica Exception

ABSTRACT Reducing high public debt is key for countries seeking to restore fiscal capacity and resilience in the wake of recent crises. But large debt reductions are rare. Jamaica stands out for reducing its debt from 144 percent of GDP to 72 percent over the last decade, a record achieved by running large, persistent primary budget surpluses. Well-designed fiscal rules combined with social partnership agreements making for fiscal ownership are at the root of its achievement.

Sharp, sustained reductions in public debt are exceptional, especially recently. We know this because public debt-to-GDP ratios have been trending up in advanced countries, emerging markets, and developing countries alike. Governments have borrowed in response to financial crises, pandemics, wars, and other emergencies, resulting in higher debt ratios. But only

Conflict of Interest Disclosure: Peter Blair Henry is on the board of Citigroup. The communications department of the International Monetary Fund (IMF) had the right to review this work prior to publication but did not inform the findings. 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. The discussant, Laura Alfaro served as a consultant for the IMF Evaluation Office and a visiting scholar at the Bank for International Settlements from 2023 to 2024. Views are the authors' and do not necessarily represent those of the IMF, its executive board, or the IMF management.

Brookings Papers on Economic Activity, Spring 2024: 133–181

rarely have they succeeded in bringing those higher debt ratios back down once the emergency has passed.

Both economic and political factors underlie this inability to reduce debt ratios. Slowing GDP growth and rising real interest rates (an unfavorable $r - g$ differential in the economist's parlance) make for adverse debt dynamics. Ideological polarization and government turnover make it hard to stay the course. Turnover creates an opportunity for a new administration to repudiate the policies of its predecessor, disrupting efforts to sustain substantial primary budget surpluses. Polarization makes it hard to agree on how to share the burden of fiscal adjustment, fraying the coalition favoring debt reduction.[1]

These economics and politics leave one pessimistic about the prospects for sustained debt reduction. Against this gloomy backdrop, it is uplifting to consider countries that have succeeded in reducing their debt ratios. In addition to their morale-building effect, such cases may help to illuminate the economic and political conditions facilitating debt consolidation.

Jamaica is such a case. The government reduced its debt from 144 percent of GDP at the end of 2012 to 72 percent in 2023. Jamaica cut its debt ratio in half despite averaging annual real growth of less than 0.75 percent over the period.[2] It did so despite vulnerability to hurricanes, floods, droughts, earthquakes, storm surges, and landslides and despite a COVID-19 pandemic that disrupted tourism and mandated exceptional increases in public spending.[3] Yet the International Monetary Fund, in its Article IV report released in 2023, forecasts a further fall in debt-to-GDP ratio to less than 60 percent over the next four years (IMF 2023).

Figure 1 shows Jamaica's achievement. It suggests that 2013 was the breakpoint when the debt ratio began its decline. Table 1 underscores the exceptional nature of the experience. Using a broad group of emerging markets and developing economies, it tabulates cases since 2000 where debt ratios fell by as much as 20, 30, or 40 percent of GDP over a five-year period. Jamaica, evidently, has few peers.

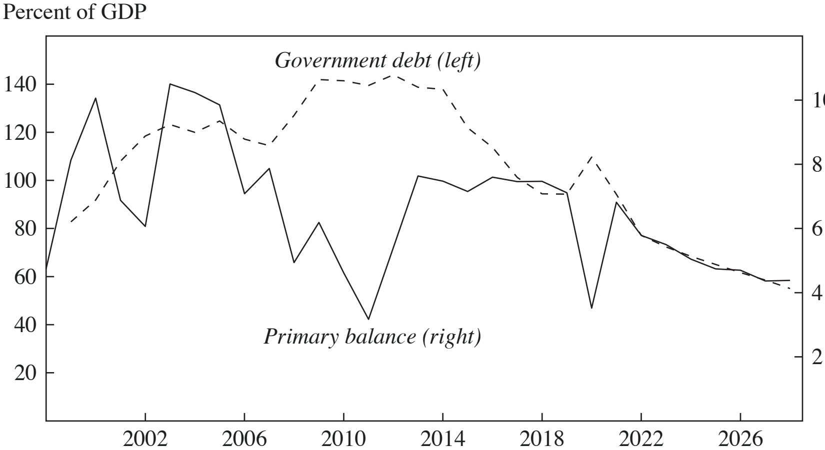
Figure 1 also points to the central economic mechanism responsible for the reduction in the debt ratio. The government ran large, sustained primary budget surpluses. Table 1 shows how unusual this is: of the debt reduction

1. Alesina and Tabellini (1990) provide a formal framework where polarization leads to overspending and debt increases, consistent with our presumption.

2. See World Economic Outlook Database (October 2023). All figures for Jamaica are for fiscal years, which run from April 1 to March 31.

3. Jamaica is ranked as the third most disaster-prone country in the world according to the Global Facility for Disaster Reduction and Recovery.

Figure 1. Jamaica: Government Debt and Fiscal Balance, 1998–2028



Source: IMF World Economic Outlook Database (October 2023).
Note: In fiscal years, which run from April 1 to March 31. Figures for 2023–2028 are projections.

episodes we identify since the turn of the century, just five relied principally on primary surpluses.[4]

The question is how Jamaica accomplished this. Our answer consists of two parts. First, Jamaica adopted fiscal rules that highlighted the debt problem, encouraged formulation of a medium-term plan, and limited fiscal slippage. The Fiscal Responsibility Framework introduced in 2010 required the minister of finance to take measures to reduce, by the end of fiscal year 2016, the fiscal balance to nil, the debt-to-GDP ratio to 100 percent, and public sector wages as a share of GDP to 9 percent (Jamaica House of Parliament 2010). The framework was augmented in 2014 to require the minister, by the end of fiscal year 2018, to specify a multiyear fiscal trajectory to bring the debt-to-GDP ratio down to 60 percent by 2026 (Jamaica House of Parliament 2014). The framework included an escape clause to be invoked in the event of large shocks. This prevented the rule from being so

4. Our paper is obviously related to the literature on fiscal consolidation, including, for example, Alesina, Perotti, and Tavares (1998); fiscal consolidation connotes episodes where governments move from large budget deficits to smaller deficits or surpluses. A difference is that in Jamaica the primary surplus already was substantial before the process of debt reduction began. We are not primarily concerned with the change in the stance of fiscal policy starting in 2013; we are focused instead on understanding a decade and more of debt reduction sustained by large, persistent primary surpluses.

Table 1. Emerging Markets and Developing Economies: Large Sustained (Five-Year) Government Debt Reductions, 2000–2022

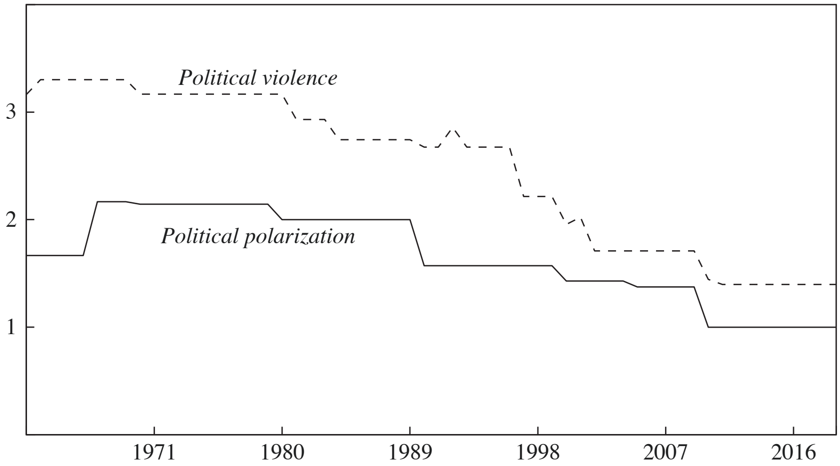
20 percent of GDP or more		*30 percent of GDP or more*		*40 percent of GDP or more*	
PAN 2005–2010	**21.2 (14.3)**	EGY 2003–2008	30.3 (–9.3)	BUL 2000–2005	42.5 (12.2)
MUS 2003–2008	21.4 (–0.4)	PRY 2002–2007	30.9 (10.9)	GEO 2002–2007	42.6 (14.9)
PHL 2003–2008	**21.4 (16.7)**	**TUR 2002–2007**	**33.7 (19.8)**	**JAM 2013–2018**	**44.4 (37.2)**
PER 2003–2008	**22.3 (15.7)**	JOR 2003–2008	34.5 (–7.3)	IDN 2000–2005	44.8 (13.7)
ARM 2002–2007	24.0 (–7.6)	BUL 2000–2005	42.5 (12.2)	LBN 2006–2011	48.9 (12.2)
EGY 2003–2008	30.3 (–9.3)	GEO 2002–2007	42.6 (14.9)		
PRY 2002–2007	30.9 (10.9)	**JAM 2013–2018**	**44.4 (37.2)**		
TUR 2002–2007	**33.7 (19.8)**	IDN 2000–2005	44.8 (13.7)		
JOR 2003–2008	34.5 (–7.3)	LBN 2006–2011	48.9 (12.2)		
BUL 2000–2005	42.5 (12.2)				
GEO 2002–2007	42.6 (14.9)				
JAM 2013–2018	**44.4 (37.2)**				
IDN 2000–2005	44.8 (13.7)				
LBN 2006–2011	48.9 (12.2)				
Average	33.1 (10.2)		39.2 (11.6)		44.6 (18.0)
N episodes	14		9		5

Source: IMF Global Debt Database.

Note: Excludes episodes associated with an external debt restructuring and major oil-exporters. The columns represent rising thresholds of debt reductions (i.e., more than 20 percent, 30 percent, and 40 percent) and the amount of debt reduction (with corresponding cumulative primary balances shown in parentheses). The countries in bold are those that reduced their debt the "old-fashioned" way (i.e., with primary balances contributing to most of the reduction). ARM = Armenia, BUL = Bulgaria, EGY = Egypt, GEO = Georgia, IDN = Indonesia, JAM = Jamaica, JOR = Jordan, LBN = Lebanon, MUS = Mauritius, PAN = Panama, PER = Peru, PHL = Philippines, PRY = Paraguay, TUR = Türkiye.

Figure 2. Jamaica: Political Polarization and Political Violence

Average rating (0-4); lower figure indicates less polarization/violence



Source: V-Dem Database (version 13).

rigid as to lack credibility. At the same time, it included clear criteria and independent oversight to prevent opportunistic use.[5]

Fiscal rules and targets do not always achieve their intended results. A quick look at the Stability and Growth Pact of the European Union (EU), which similarly targets a 60 percent debt-to-GDP ratio, is a stark reminder of this fact.[6] This brings us to the second part of our answer: elected officials leveraged Jamaica's hard-won tradition of consensus building—of constructing over the course of thirty years social partnerships aimed at facilitating dialogue, limiting political instability, and reducing political polarization and violence (see figure 2). In 2013, a series of ongoing discussions in the National Partnership Council (NPC), a social dialogue collaboration involving the government, parliamentary opposition, and social partners, culminated in the Partnership for Jamaica Agreement on consensus policies in four areas, first of which was fiscal reform and consolidation. The Partnership for Jamaica Agreement fostered a common belief that the burden of adjustment would

5. Jamaica is unusual in this regard; it is one of only two Caribbean Community (CARICOM) countries, along with Grenada, to have adopted an explicit national fiscal rule. Grenada's national budget balance, debt, and expenditure rules date from 2015—that is, they postdate Jamaica's fiscal rule.

6. European Commission, "Stability and Growth Pact," https://economy-finance.ec.europa.eu/economic-and-fiscal-governance/stability-and-growth-pact_en.

be widely and fairly shared. It supported the creation of the Economic Programme Oversight Committee (EPOC) to monitor and report on fiscal policies and outcomes, providing independent verification that all parties kept to the terms of their agreement.

EPOC and the Partnership for Jamaica Agreement solidified the sharp decline in political polarization that began four years earlier, coincident with creation of the NPC.[7] Less polarization made for policy continuity when a different political party took power in 2016. For the first time in decades, a new government did not reverse the fiscal policies of its predecessor. By creating a sense of fair burden sharing, Jamaica's organized process of consultation sustained public support for the country's fiscal rules, culminating in March 2023 with the establishment of a permanent, independent fiscal commission.

As always, the full story is more complex. Jamaica managed its financial system well. It adeptly managed the term structure of the debt. But the two elements highlighted above—a well-designed fiscal rule and a partnership agreement creating confidence that the burden of adjustment would be widely and fairly shared—were key. Neither element would have worked in the absence of the other. Both were needed.

An important question is whether the lessons from Jamaica generalize. We discuss two other countries that achieved significant debt reduction by adopting fiscal rules and consensus-building arrangements: Ireland in the late 1980s and Barbados for a decade starting in the early 1990s. These cases differ in their particulars. But they have in common that Ireland and Barbados—like Jamaica—are small, open economies. These economies are highly structured, in that trade unions and employers' associations are cohesive and powerful. In both cases, agreements were reached and institutions were created to initiate and maintain the momentum of debt reduction, leveraging earlier historical experience with institution-based consensus building.

These similarities are consistent with the literature suggesting that democratic corporatism, a process of policy formulation involving extensive consultation and consensus building, is the easiest where interest groups are well-organized and the number of agents is limited.[8] They are consistent

7. Cause and effect are admittedly difficult to disentangle in this context. It is reasonable to believe that causality ran both ways. We return to this issue in section III.D below.

8. Peter Katzenstein, who popularized the concept of democratic corporatism, defines it as a political system characterized by "an ideology of social partnership, a centralized and concentrated system of economic interest groups, and an uninterrupted process of bargaining

with the view that such arrangements are imperative in small, open economies exposed to external shocks. And they are consistent with the view that so-called neo-corporatist arrangements, when and where they emerge, build on earlier historical experience.

I. Historical Background

Jamaica's recent experience of debt reduction is exceptional, but the country's earlier history was also marked by exceptional fiscal developments, some positive, others not. The 1962 constitution included a provision prohibiting the government from borrowing without parliamentary approval. It prioritized servicing the debt as an obligation senior to other government expenses (Langrin 2013). Accordingly, Jamaica has never had an outright default on its sovereign debt, although it has conducted domestic debt exchanges (described below). Fiscal restraint was designed to attract the foreign direct investment (FDI) needed for development of the capital-intensive bauxite industry. True to form, FDI financed 30 percent of all capital formation in the 1960s and virtually all investment was in the bauxite sector.

Public debt remained modest in the first post-independence decade, reflecting the consensus around these priorities. The ruling Jamaica Labour Party (JLP) eschewed activist fiscal and monetary policies, relying on tax breaks and free profit repatriation to attract foreign investment.[9] Jamaica successfully grew the denominator of the debt-to-GDP ratio: real GDP rose by roughly 6 percent per annum in what Stone and Wellisz (1993, 140)

among all of the major political actors across different sectors of policy" (Katzenstein 1985, 80). We are not arguing that democratic corporatism is the only setting in which significant debt reduction can occur. One can think of authoritarian settings where high debts were dramatically reduced; Romania under Nicolae Ceauşescu springs to mind (not that this turned out well for the Ceauşescus). Two of the fourteen cases in table 1 have a rating of 0.4 or below on the polity scale, situating them on the autocratic side of the autocracy-democracy continuum. Others have relatively high levels of political polarization but were able to reduce debt through other means (high inflation, financial repression, or faster economic growth). But to reiterate, our goal here is not to determine whether democracy or autocracy is "better" for debt reduction. It is to understand how Jamaica did it.

9. Thus, fiscal deficits averaged a relatively modest 2.3 percent of GDP from 1962 through 1972 (Henry and Miller 2009), while the currency was pegged to the pound sterling under a quasi-currency board system. Jamaica switched from a sterling peg to a dollar peg in 1973, following the change in government (which reinforced the peg with capital controls—more on which below), the United States by this time having become the country's leading trade partner.

called “one of the best growth records in the world.” Mining was relatively unimportant in the 1950s, and tourism had contributed only modestly to economic activity; this meant that there was low-hanging fruit to be picked. King (2001, 7) describes growth in this period as built on “natural endowments of bauxite and beaches.”

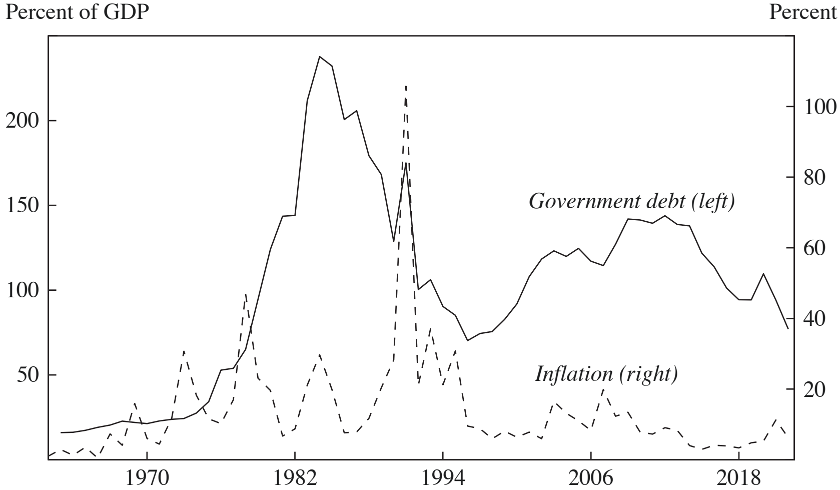
Capital-intensive mining created little employment, however, while Dutch disease pressures led to declines in the relative importance of agriculture, forestry, and fishing. Small-scale manufacturing and services had limited capacity to absorb surplus labor released by the rural sector, given the floor placed on wages by strong unions and insider-outsider dynamics.

By the time of the 1972 election, unemployment, mostly urban, had risen to more than 20 percent, and dissatisfaction with education and health care services was rife. These grievances led to a backlash against the JLP’s cautious policies, culminating in the electoral victory of the People’s National Party (PNP) led by the charismatic Michael Manley. The approach of the new PNP government was variously labeled “state populism” and “democratic socialism.”[10] The PNP nationalized companies, raised import barriers, and imposed exchange controls; spending on schooling, food subsidies, and public housing exploded (Henry 2013). Public employment rose by two-thirds between 1972 and 1977, while public spending as a share of GDP doubled from 23 percent to 45 percent. The budget deficit averaged 15 percent of GDP. The government financed what it could by borrowing, and the Bank of Jamaica financed the rest. The debt-to-GDP ratio soared from 24 percent at the time of the 1972 election to 124 percent in 1980 (figure 3). Inflation, having averaged 4 percent in the first post-independence decade, reached 27 percent in 1980.

The PNP’s focus on social justice notwithstanding, its policies were economically disastrous. Dirigiste rhetoric and policies of nationalization discouraged investment. Labor productivity and real wages plummeted, and unemployment rose to 27 percent in 1980 (Henry 2023). As standards of living continued to fall, the implications for survival of the zero-sum patronage gained or lost with each election rose higher, and political violence spiked (figure 2). This economic and political chaos led, predictably, to the PNP’s defeat in the 1980 election, the return of the JLP, and a swing back toward more market-oriented policies.

10. The party used the latter term in its election manifestos, as do Stephens and Stephens (1986) in their analysis of Jamaican political economy.

Figure 3. Jamaica: Government Debt and Inflation, 1962–2022

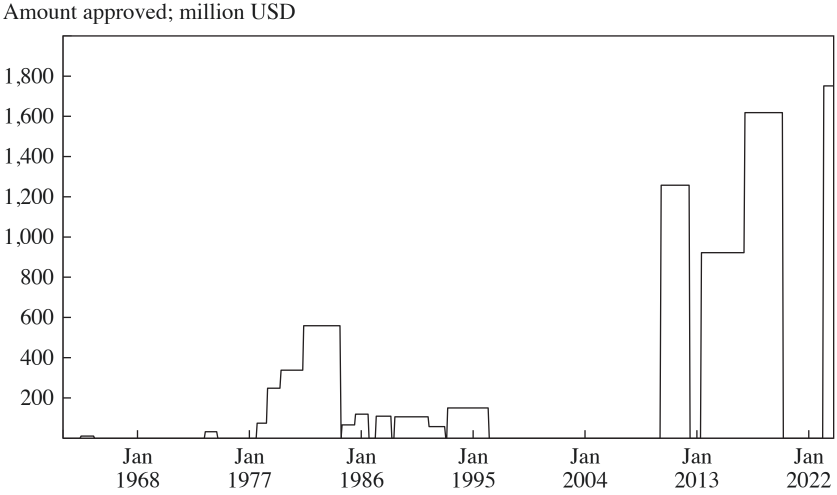


Source: IMF data from the Global Debt Database, International Financial Statistics, and the World Economic Outlook Database (October 2023).
Note: In fiscal years, which run from April 1 to March 31. Inflation is as of end of period.

When the decline in foreign investment and macroeconomic stimulus created balance of payments problems in 1977–1978, the PNP was forced to negotiate agreements with the IMF. Both programs were then suspended when the government failed to meet performance targets. (Figure 4 shows a timeline of the country's agreements with the IMF.) The JLP government tried again in 1980: it devalued to improve export competitiveness, cut government spending, eliminated price controls, and negotiated new loans with the IMF and World Bank (Kirton and Ferguson 1992). Its policies were contractionary in the short run, provoking violent demonstrations, but by the mid-1980s, productivity and GDP began rising again.

Jamaica's first episode of debt reduction then began in the second half of the 1980s. Debt had risen to an extraordinarily high 240 percent of GDP, requiring urgent action. The JLP imposed spending cuts, moving the primary balance into surplus. Progress was interrupted in 1988–1989 by Hurricane Gilbert, which destroyed more than 100,000 homes, but even this did not throw the process off course. Importantly, when the PNP returned to power in 1989, it maintained the same basic economic stance. Chastened by its earlier experience of deficits and negative growth, it restrained public spending, raised taxes, and restricted credit, allowing primary surpluses to be

Figure 4. Jamaica: Financial Arrangements with the IMF



Source: IMF.

maintained. There was now more dialogue between the parties as their policy differences grew less pronounced. Figure 2 shows the measure of political polarization from the Varieties of Democracy (V-Dem) Database; this is based on responses to a survey of country experts conducted each year.[11] The figure documents a fall in polarization at the beginning of the 1990s, the largest fall since independence, exceeding even the sharp fall in polarization two decades later. This was then followed by a steep decline in political violence from the mid-1990s to early 2000s (also shown in figure 2). This experience thus shows how cross-party agreement on basic economic priorities is important for debt reduction.

This is the positive part of the story. The negative part is inflation, which was the single most important contributor to debt reduction in the decade ending in 1995. End of fiscal year inflation accelerated to 28 percent in 1990,

11. V-Dem (https://www.v-dem.net) takes input from at least five experts for each country and year, drawing on 3,700 experts worldwide. For political polarization, it asks: "Is society polarized into antagonistic, political camps?" Responses range from: 0 (Not at all. Supporters of opposing political camps generally interact in a friendly manner); to 4 (Yes, to a large extent. Supporters of opposing political camps generally interact in a hostile manner). For political violence, it asks: "How often have non-state actors used political violence against persons this year?" Responses range from: 0 (Not at all. Non-state actors did not use political violence); to 4 (Often. Non-state actors often used political violence).

106 percent in 1991, and 21 percent again in 1992. Given that some 60 percent of local government debt was held in medium- to long-term securities, this brought the debt ratio down very sharply, from 175 percent of GDP at the end of 1990 to 100 percent at the end of 1991, for example (IMF 2000).

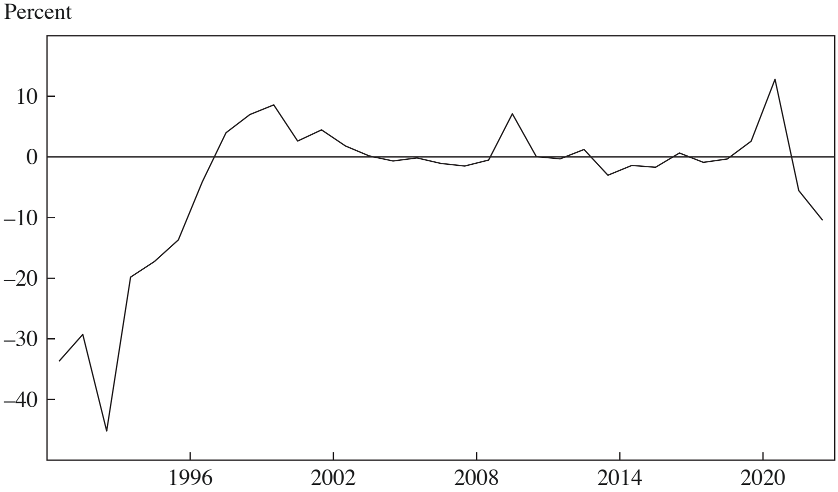
But this route to debt reduction was unsustainable because it undermined the foundations of the financial system. Inflation reflected measures taken by the Jamaican authorities to liberalize the financial system, without at the same time strengthening financial supervision. In the run-up to the crisis, they removed ceilings on credit provided by banks, deregulated deposit rates, encouraging banks to compete for funding, and permitted banks to make US dollar denominated loans (Kirkpatrick and Tennant 2002).[12] Unfortunately, even while Jamaica began easing restrictions on capital account transactions, it retained a patchwork of financial regulators and regulations, creating scope for regulatory arbitrage given the weak supervisory capacity of the central bank. The authorities liberalized the capital account in the hope that this would lead to capital inflows, reduce depreciation pressure on the exchange rate, and mitigate inflation. This was also when the IMF was advising its emerging market members to liberalize the capital account, and Jamaica, continuously under IMF programs, acted accordingly.

The result was a very large capital inflow as exchange controls were relaxed, funding additional domestic lending as investors repatriated offshore dollars. The removal of quantitative credit ceilings permitted the development of an enormous credit boom; bank credit to the private sector grew at double-digit rates, always a warning sign, hence the surge of inflation. The credit boom was characterized by deteriorating asset quality, declining bank profits, and a growing currency mismatch as banks extended US dollar loans to firms in the nontraded goods sector where revenues accrued in local currency.

Initially, the implications for the debt-to-GDP ratio were favorable, as the credit-fueled burst of inflation led to a negative real interest rate/real growth rate differential (figure 5). But those favorable dynamics did not last. In mid-1995, the Bank of Jamaica finally got serious about inflation and tightened monetary policy. Higher interest rates led to weakness in the real estate

12. There had in fact been an earlier attempt to liberalize the banking system in the mid- to late 1980s as a condition of the country's World Bank program, but this was reversed in 1989 when Hurricane Gilbert prompted sharp increases in government spending, which the fiscal authorities enlisted the banks to finance. Another factor prompting reregulation was a massive inflow of reinsurance funds, leading to increased bank liquidity and what was perceived as an unsustainable surge in lending.

Figure 5. Jamaica: $r - g$ Differential, 1990–2022



Source: IMF Global Debt Database and World Economic Outlook Database (October 2023).
Note: r is calculated as the effective interest rate on government debt deflated by the GDP deflator. g is the real GDP growth rate.

sector, to which financial institutions were predictably committed. This raised questions about bank solvency, precipitating withdrawals by panicked depositors.[13] A massive financial crisis engulfed commercial banks, investment banks, building societies, insurance companies, and security brokers in the mid-1990s. Laeven and Valencia (2020) rank this as the third most costly banking crisis anywhere in the world in the five decades after 1970.

Starting in 1996, GDP fell for three consecutive years.[14] With nonperforming loans as a share of total loans rising to nearly 30 percent, the financial system had to be recapitalized by a special purpose vehicle, the Financial Sector Adjustment Company, whose liabilities were ultimately transferred to the government's balance sheet. Effectively, the government replaced nonperforming loans with government debt in an effort to reassure depositors.

13. Newly deregulated life insurance companies aggressively marketed short-term products offering high rates of return and invested these short-term funds in long-term, illiquid assets, mainly real estate. Scenting an opportunity, banks for their part extended high interest rate loans to insurance companies with which they were connected, causing the banking system to be implicated. This is a clear instance of the regulatory arbitrage noted above.

14. See IMF Global Debt Database and World Economic Outlook Database.

Given a fiscal cost of 44 percent of GDP and falling revenues owing to the crisis-induced recession, it is no surprise that this mega-financial crisis threw debt reduction off course. After falling steadily for more than a decade, the debt ratio now rose sharply. This episode is a reminder that financial stability is essential for sustained debt reduction, and that a burst of inflation, even if helpful for debt reduction in the short run, is not compatible with such stability.

The debt ratio continued rising through the first decade of the new century, approaching 150 percent of GDP in 2010. It did so even once the central government resumed running primary surpluses. About half the increase in the debt ratio in the 2006–2011 period was due to currency depreciation that raised the real burden of foreign currency debt and an unfavorable real interest rate/real growth rate differential, reflecting anemic growth together with stubbornly high nominal interest rates in the range of 15 percent.[15] The other half was due to the deficits of public bodies, such as the Urban Development Corporation and Bauxite and Alumina Trading Company, of which there were more than two hundred in number, and debt to the Venezuelan state-owned oil company Petróleos de Venezuela (PDVSA), which was incurred by Petrojam, a limited liability company, but guaranteed by the Jamaican government.[16]

II. What Jamaica Did

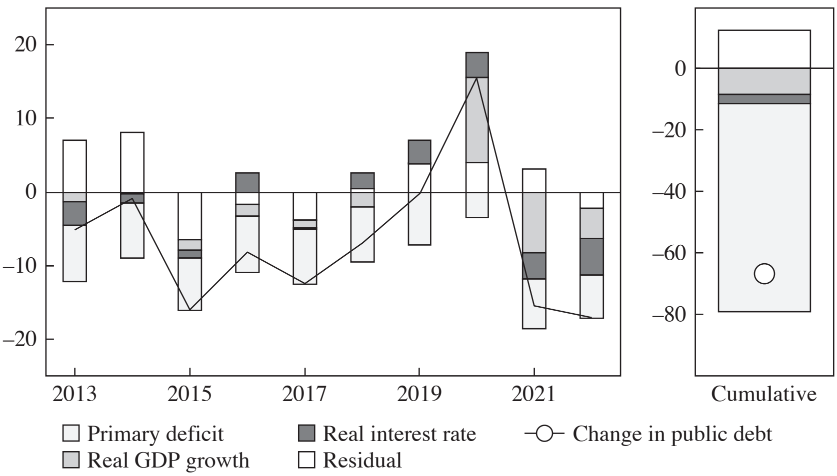
This unpropitious backdrop renders what happened next all the more remarkable. As shown in figure 1, the debt-to-GDP ratio stopped rising in 2010 and, after a few years of relative stability, began falling precipitously, from 144 percent in 2012 to just 72 percent in 2023.[17] This achievement is

15. The root causes of this slow growth were several. Jamaica lacked affordable energy to refine bauxite into aluminum and inexpensive labor to compete with low-cost Caribbean and Central American neighbors. Infrastructure, education, and training were deficient. High real interest rates for their part reflected chronic doubts about the government's willingness and ability to control inflation and service its debts.

16. At the time, Petrojam was owned jointly by the Petroleum Corporation of Jamaica, an entity of the Jamaican government, and PDVSA. Conventional debt accounting, as in equations (1) and (2) below, includes these two items in the residual contributing to changes in the debt ratio rather than subtracting them from the primary balance. If one instead subtracts them from the primary balance, primary surpluses in the 2006–2011 period become less impressive (they fall from an average of 5.6 percent of GDP to 1.9 percent of GDP). But this does not change the fact that the primary balance was already in surplus. We return to this below.

17. The IMF expects that debt ratio to decline still further, to below 60 percent four years from now.

Figure 6. Jamaica: Drivers of Debt-GDP Dynamics, 2013–2022



Source: Authors' calculations.
Note: In fiscal years, which run from April 1 to March 31.

exceptional (in several senses of the word). We first analyze how this debt reduction was achieved in an accounting sense, before asking how it was achieved in an economic and political sense.

To this end, figure 6 shows the standard debt decomposition:

$$\Delta b = d + \frac{(r-g)}{1+g} b_{t-1} + sfa, \tag{1}$$

where b is debt as a share of GDP and Δb is its change. The right-hand side is made up of the primary budget deficit (net of interest payments) relative to GDP, denoted d; $r-$ g interacted with the inherited debt ratio; and the residual, which captures defaults, restructurings, conversions, assumption by the public sector of private debt, other off-budget spending, and exchange rate effects, collectively denoted *sfa* (stock-flow adjustment).

Figure 6 shows that debt reduction was driven mainly by primary budget surpluses, which are large throughout the period (excepting only 2020, the first year of COVID-19). Existing primary surpluses were raised by an additional 2 percentage points of GDP in fiscal years 2012 and 2013, mainly through expenditure cuts as a share of GDP (see table 2).[18] Subsequently, the government maintained these primary surpluses despite strongly increasing noninterest spending, from 19 percent of GDP in 2014 to 24 percent of GDP in 2019, on the eve of the COVID-19 crisis.[19] Following the initial spending adjustment, in other words, surpluses were sustained by strongly increasing tax revenues as a share of GDP.[20] Most of these gains in revenue resulted from broadening the tax base (removing exemptions), although in addition, there were an increase in the personal income tax rate for high earners and improvements in tax administration.

There was also a modest contribution from GDP growth, mainly toward the end of the period, modest because growth remained anemic. This is a reminder that sound debt management is no guarantee of positive growth performance—and, conversely, that strong growth is not always and everywhere a prerequisite for successful debt reduction.[21]

Might the large primary surpluses needed for debt reduction have themselves slowed growth? Hypothetically, by pushing back the deadline for reaching a 60 percent debt-to-GDP ratio or raising that target, the government might have undertaken more social spending, boosting aggregate demand. Of course, to infer the impact on growth one would need a fully specified model of the Jamaican economy, robust to policy regime. In any case, such growth of output as occurred reflected strong increases in employment, not increases in productivity, consistent with the idea that problems of

18. The decline in spending was spread across capital projects, central government purchases of goods and services, and (to a much smaller extent) the public sector wage bill. The concurrent increase in revenues reflected transfers from the National Housing Trust (which makes low interest rate loans to housing developers) and renewal of licenses by two telecom companies. These were one-off receipts rather than structural revenue measures, in other words. Details are given in Government of Jamaica (2013).

19. In addition, there was a trend decline in interest spending as a share of GDP, as table 2 also shows, reflecting the falling debt ratio and a trend decline in sovereign spreads as Jamaica's fiscal position strengthened.

20. Thus, Jamaica does not fall neatly into either the spending-reduction or tax-increase categories distinguished by Alesina, Perotti, and Tavares (1998). From the long-run perspective of concern to us here (since we are focused on how primary surpluses were sustained over a decade and more), this was a revenue-driven consolidation, a member of the category that Alesina, Perotti, and Tavares (1998) question can be sustained. Thus, our conclusions here contrast with theirs.

21. There is also a contribution to debt reduction from the negative real interest rate, reflecting high inflation in the immediate post-COVID-19 period. Otherwise, real interest rates are modestly positive on average (figure 5), roughly offsetting the contribution of real GDP growth (figure 6). There is a sharp fall in both inflation and nominal interest rates on government debt in 2014, leaving the real interest rate essentially unchanged.

Table 2. Jamaica: Summary of Government Operations (Percent of GDP)

	FY09	*FY10*	*FY11*	*FY12*	*FY13*	*FY14*	*FY15*	*FY16*	*FY17*	*FY18*	*FY19*	*FY20*	*FY21*	*FY22*
Overall balance	–11.1	–6.3	–6.4	–4.1	0.1	–0.5	–0.3	–0.2	0.5	1.2	0.9	–3.1	0.9	0.3
Revenues	27.5	26.8	25.6	25.7	27.1	26.2	27.0	28.0	29.1	30.6	30.6	29.5	31.0	30.1
Expenditures	38.7	33.2	32.0	29.8	27.0	26.7	27.3	28.2	28.6	29.4	29.7	32.6	30.1	29.8
ow/Interest expense	17.3	11.0	9.6	9.5	7.5	8.0	7.4	7.8	7.0	6.3	6.2	6.6	5.9	5.5
Primary balance	6.2	4.6	3.2	5.4	7.6	7.5	7.2	7.6	7.5	7.5	7.1	3.5	6.8	5.8
(Change from previous year in percentage points of GDP)														
Overall balance	—	4.8	–0.1	2.3	4.2	–0.6	0.2	0.1	0.6	0.7	–0.3	–4.0	4.0	–0.6
Revenues	—	–0.7	–1.3	0.2	1.4	–0.9	0.7	1.0	1.1	1.5	0.0	–1.1	1.5	–0.9
Expenditures	—	–5.5	–1.2	–2.2	–2.8	–0.3	0.5	0.9	0.5	0.8	0.3	2.9	–2.5	–0.3
ow/Interest expense	—	–6.4	–1.4	–0.1	–2.0	0.5	–0.5	0.4	–0.8	–0.7	–0.1	0.4	–0.7	–0.4
Primary balance	—	–1.6	–1.5	2.2	2.2	–0.2	–0.3	0.4	–0.1	0.0	–0.4	–3.6	3.3	–1.0

Source: IMF World Economic Outlook Database (October 2023).

education and training, and not inadequate demand, are at the root of slow growth.[22] In addition, the government's debt reduction strategy also reduced the volatility of growth; prior to COVID-19, Jamaica had nineteen consecutive quarters of growth, where the longest earlier span was nine quarters.[23]

Figure 6 highlights several years early in the period in which there were increases in the debt burden due to factors not otherwise explained. These increases reflect the materialization of contingent liabilities stemming from unexpected losses by public enterprises such as Clarendon Alumina Production (CAP) and Jamaica Urban Transit Company. In fiscal year 2012, for example, the government was forced to assume 70 percent of the liabilities of CAP (IMF 2018). The prevalence of such problems was then reduced in the period's second half by strengthened governance of public enterprises (as we explain in section III.A below). Meanwhile, stronger financial supervision and regulation helped to avoid losses from the kind of banking crisis that had thrown 1990s debt-reduction efforts off course.

Figure 7 sheds more light on what lies behind the debt decomposition. Here we further decompose the change in the debt-to-GDP ratio as follows:

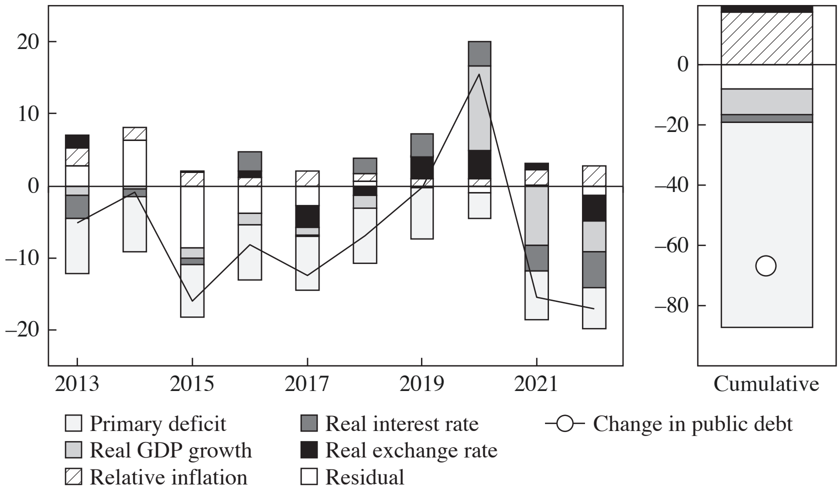
$$\Delta b = d + \frac{(r-g)}{1+g} b_{t-1} + \frac{z\,a}{(1+g)(1+p^*)} b_{t-1} + \frac{(p-p^*)a}{(1+g)(1+p)(1+p^*)} b_{t-1} + sfa \qquad (2)$$

where r = the real interest rate; p = growth rate of GDP deflator; p^* = growth rate of US GDP deflator; g = real GDP growth rate; a = share of foreign currency denominated debt; z = real exchange rate depreciation (measured as $[(e_t/e_{t-1})(1 + p^*)/(1 + p)] - 1$); and e = nominal exchange rate (measured by the local currency value of the US dollar). The exchange rate matters because more than a quarter of debt at the beginning of the debt reduction period was denominated in or indexed to dollars. Comparing figures 6 and 7, we can see how ongoing depreciation of the Jamaican dollar increased the domestic currency value of external debt.

22. See also footnote 15 on these problems. Kandil and others (2014) noted prior to the debt-reduction episode that Jamaica had the highest elasticity of employment with respect to output in the Caribbean.

23. Statistical Institute of Jamaica, "National Accounts," https://statinja.gov.jm/NationalAccounting/nationalaccountsnotes.aspx.

Figure 7. Jamaica: Drivers of Debt-GDP Dynamics Accounting for Real Exchange Rate and Relative Inflation Effects, 2013–2022



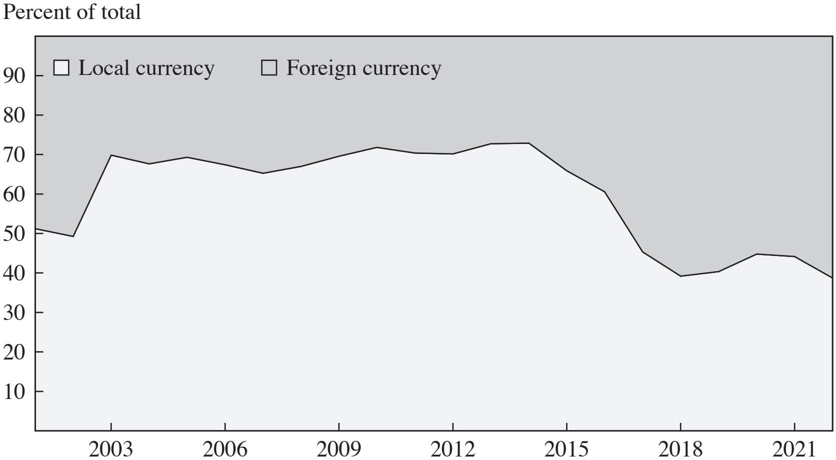
Source: Authors' calculations.
Note: In fiscal years, which run from April 1 to March 31.

Relatedly, figure 8 shows how the foreign currency share of the debt rose as debt reduction allowed Jamaica to resume tapping international financial markets in 2014. It didn't hurt that this was a period of low interest rates in advanced countries, encouraging international investors to search for yield in emerging markets. While a limited number of relatively large middle-income countries were able to place domestic currency debt with international investors over this period (freeing themselves of the "original sin" of foreign currency denominated external debt), Jamaica was not one of these.[24]

The country's increasing reliance on foreign currency debt was not overly detrimental. Figure 7 shows why: although there was a contribution to the debt from exchange rate depreciation, the real exchange rate was reasonably stable against the US dollar (that is, the nominal exchange rate moved

24. Arslanalp and Eichengreen (2023) show how success at placing domestic currency denominated securities with international investors has been largely limited to a handful of relatively large emerging market economies. In November 2023, Jamaica issued its first-ever Jamaican dollar-linked bond in international capital markets, with "the [Government of Jamaica's] objective of opening local currency debt issues to international investors" (Ministry of Finance and the Public Service 2023, par. 2). Jamaica used the proceeds to buy back outstanding US dollar denominated bonds, which Moody's commented would reduce "the government's exposure to foreign exchange risk, which is a credit positive" (ibid., par. 4).

Figure 8. Jamaica: Currency Composition of Government Debt



Source: IMF.
Note: In fiscal years, which run from April 1 to March 31.

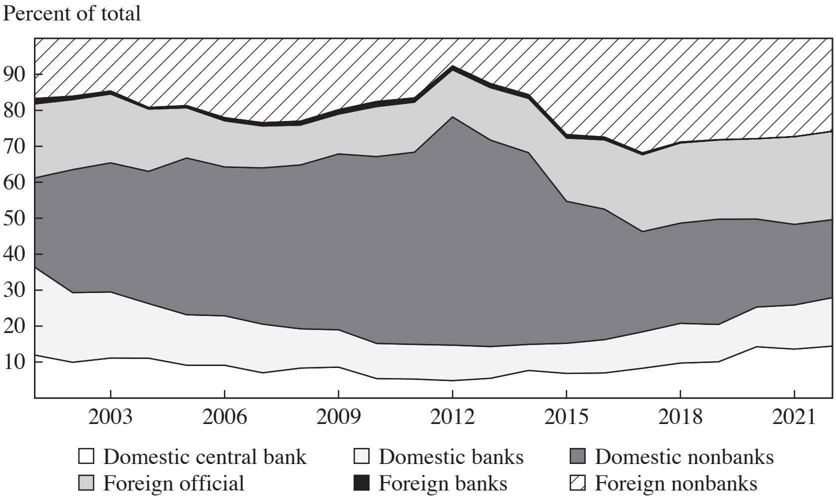
broadly in line with the inflation differential vis-à-vis the United States). This is a reminder of the value of a relatively stable real exchange rate for debt reduction, especially when a portion of that debt is denominated in foreign currency.[25]

Comparing figures 6 and 7, we see that separating out the impact of exchange rate depreciation on the value of external debt turns the overall contribution of the *sfa* from positive (adding to the debt burden) to negative (subtracting from the debt burden).[26] This makes it tempting to look to the

25. This is not to recommend issuing debt in foreign currency to take advantage of relatively low international interest rates. The risks are well known. The strategy worked in Jamaica because the authorities limited real depreciation of their currency. The credibility of Jamaica's policies, discussed further below, may help to explain the stability of the currency in the face of global shocks. So too may an element of luck.

26. Figures 6 and 7 show that at least as important as the 2013 debt exchange were financial operations undertaken in 2015 and 2016. The 2015 residual reflects a buyback at substantial discount of the government's Petrocaribe debt. In 2015 the government bought back this debt from cash-strapped Venezuela, raising cash by issuing a thirteen-year Eurodollar bond. The buyback replaced debt to Venezuela with new external debt bearing a lower face value but a higher interest rate. The net effect was to push a portion of the financial burden out into the future, creating a 10 percent of GDP reduction in measured debt in 2015 (Okwuokei and van Selm 2017). The 2016 residual reflects an accounting adjustment implemented in conjunction with the new fiscal responsibility law, described below, that excluded intragovernmental debt holdings and the Bank of Jamaica's external debt (offset by the central bank's external reserves), in line with international statistical standards.

Figure 9. Jamaica: Holders of Government Debt, 2001–2022



Source: Updated from Arslanalp and Tsuda (2014).

pair of domestic debt restructurings conducted in 2010 and 2013. In fact, these operations had a limited impact on the debt burden. Neither entailed nominal haircuts reducing the face value of the debt, partly because a non-negligible fraction of that debt was held by domestic financial institutions (figure 9) whose stability would have been jeopardized (Schmid 2016). External debt was excluded because Jamaica's global bonds lacked majority action clauses, threatening litigation and inconclusive negotiations with holdout creditors.[27]

Still, these exchanges helped on the budgetary front, despite the absence of face-value haircuts, by reducing coupons and extending maturities. In both cases, the government succeeded in achieving very close to 100 percent investor participation (Langrin 2013). Here the same factor that prevented face-value haircuts—that domestic debt was held mainly by a handful of

27. Such clauses allow a qualified majority of creditors to cram down restructuring terms on a dissenting minority. The exclusion of external debt from restructuring was a factor in the government's ability to tap the Eurodollar market in 2014 and 2015 and buy back the Petrocaribe bonds, as described in the preceding footnote. In addition, the constitutional priority attached to servicing external debt put in place in the 1960s to help attract FDI may have contributed to the difficulty of restructuring.

financial institutions—helped by attenuating free-rider problems.[28] This observation has implications for whether the lessons from Jamaica carry over to other countries, since in quite a few other countries debt securities are not in the hands of domestic banks but are widely held by heterogeneous creditors whose coordination is difficult to achieve.

In sum, the Jamaican authorities mainly reduced their debt "the old-fashioned way," by running substantial primary surpluses for an extended period. To be sure, they also grew the economy, if modestly, while eschewing excessive currency depreciation that might have elevated the domestic currency value of external debt. They avoided financial instability that had caused the materialization of contingent liabilities and derailed earlier efforts at debt reduction. They engaged in some clever financial management. But budget surpluses were key.

This strategy of running substantial primary budget surpluses for extended periods is not commonplace; other emerging markets, developing countries, and advanced economies would be envious. The question is how Jamaica did it.

III. How Jamaica Did It

Our explanation of how Jamaica did it consists of two parts. First, Parliament passed a set of rules known as the Fiscal Responsibility Framework. These rules highlighted the debt problem, legislated formulation of a medium-term plan, and made it easier to define and detect fiscal slippage.

All too often, however, rules are honored in the breach. This brings us to the second element: Jamaica leveraged its hard-won tradition of forging social partnerships to establish consultative bodies with the legitimacy, independence, and stature needed to build and sustain a social consensus for fiscal adjustment, while credibly monitoring and reporting on the government's adherence to its fiscal rules and the progress of the overall economic

28. Thus, the government could coordinate its negotiations with this limited number of financial institutions over which it had regulatory oversight. As an inducement, financial institutions that participated received preferential access to a Financial Sector Support Fund administered by the central bank. Participants in the 2010 debt exchange had the option of new series that were CPI indexed and noncallable, features not included in the old bonds. In the 2013 exchange, large institutional investors that initially held out were subjected to political pressure (they were criticized as "unpatriotic"), while small retail investors who might have held out from a second restructuring that further lengthened maturities were offered special one-year bonds.

reform program. In 2009, government, the opposition, business, trade unions, and civil society groups formed a consultative body called the National Partnership Council (NPC) to address the effects of the global financial crisis as well as long-standing economic and social issues. Deliberations of this council enabled stakeholders to exchange views, provide input, reach consensus about the societal importance of debt reduction, and assure all partners that the burden of adjustment would be broadly and fairly shared. In ongoing meetings, its members discussed the conformity of policies with their shared priorities and suggested changes to align policies and priorities more closely. The Fiscal Responsibility Framework we discuss in the next subsection can be seen as a legislative response to the broad societal consensus for fiscal restraint built by the NPC. It then became possible to move from vision to reality when EPOC was created in 2013. EPOC consists of representatives of the private and public sectors, unions, and civil society but with disproportionate representation of the financial sector. It is tasked specifically with monitoring the government's progress and benchmarking this against the performance targets of the Fiscal Responsibility Framework.

This monitoring, dialogue, and consensus building were pivotal for holding government accountable for its budgetary actions and for maintaining the consensus needed to get the process on track and keep it there.

III.A. Fiscal Rules

Prior to 2010, when the Fiscal Responsibility Framework was put in place, Jamaica's best-laid fiscal plans repeatedly went off course. Recorded deficits exceeded those in the budget passed by Parliament in every year between 2003 and 2009.[29] Growth forecasts were excessively rosy. Tax revenues regularly fell short of projections. Expenditure overshot what was budgeted; in particular, public sector wage settlements regularly exceeded what was assumed by the Ministry of Finance. Public entities did not regularly report to the Ministry of Finance. For its part, the ministry did not update cash flow forecasts and performance for these entities in-year, unlike for the central government. Lack of updating permitted chronic overspending and the accumulation of arrears by these public bodies. (We described in section II how the deficits of public entities such as CAP and Jamaica Urban Transit contributed to the growth of debt.) The central government conducted budgeting on a year-by-year basis; "the future implications of expenditure

29. Ministry of Finance and the Public Service, Government of Jamaica, "Ministry Papers," https://www.mof.gov.jm/ministry-papers/.

decisions [were] not elaborated on in the budget documents . . . consideration is not always given for the medium/long-term implications of decisions made in the short-term" (Leon and Smith 2012, 14). Though the Ministry of Finance was responsible for describing its debt management strategy in broad terms, it was not required to formulate and present a debt sustainability analysis.

The 2010 Fiscal Responsibility Framework, formally an amendment to existing financial administration and audit regulations, addressed most of these shortcomings.[30] It anchored budgeting by requiring the minister of finance to take appropriate measures to reduce, by the end of fiscal year 2016: (a) the fiscal balance to nil; (b) the ratio of debt to GDP to 100 percent; and (c) public sector wages as a share of GDP to 9 percent (Jamaica House of Parliament 2010). The framework was tightened in 2014 to require the minister, by the end of fiscal year 2018, to specify a multiyear fiscal trajectory bringing the debt-to-GDP ratio down to no more than 60 percent by fiscal year 2026 (Jamaica House of Parliament 2014).[31]

Importantly, these numerical targets for debts and deficits came with an escape clause to be invoked in exceptional circumstances. Rigid targets would have lacked credibility in an environment prone to hurricanes, floods, and other natural disasters; the government's assertion that under no circumstances would it respond to such events with a revised budget would not have been taken at face value. At the same time, an escape clause not limited to events beyond control of the government, lacking explicit thresholds for activation and with no provision for independent verification, would have been destabilizing; it would have given the government free rein to disregard its targets. As Valencia, Ulloa-Suárez, and Guerra (2024) describe, a well-designed escape clause must be accompanied by clear triggers and conditions, clear assignment of responsibility for activation and deactivation, and a clear communication strategy.

30. The IMF and World Bank made adoption of the Fiscal Responsibility Framework a condition of their 2010 lending programs but were unhappy with the incomplete rules adopted that year; no immediate changes were made, since the IMF agreement went offtrack almost immediately, and disbursements were halted. The IMF then required strengthening of the framework as a condition for its 2013 arrangement, and the amendments followed in 2014.

31. The 2026 deadline was pushed back to 2028 due to the pandemic, in an example of the operation of the escape clause mechanism described below. The rationale for the separate public sector wage target was that wage compensation was a principal driver of the fiscal balance. Subsequent experience showed that even when the wage target was missed it still could be possible to meet the debt and deficit targets; correspondingly, the separate wage target was eliminated in 2023.

Jamaica's escape clause satisfies these prerequisites. It can be activated only in response to a natural disaster, a public health or other emergency, or a severe economic downturn (of 2 percent of GDP in a quarter). It can be invoked only after verification by the auditor general, whose independence from other government agencies is guaranteed by the constitution, that the fiscal impact exceeds a minimum threshold of 1.5 percent of GDP. The auditor general must submit its assessment to Parliament, along with supportive documentation from the Ministry of Finance, and suspension of the fiscal rule must be approved by both Houses. Valencia, Ulloa-Suárez, and Guerra (2024) rate escape clause clarity on six dimensions and give Jamaica's escape clause a rating well above the Latin American and Caribbean average.

The government was thus able to invoke this escape clause in response to COVID-19, reducing the VAT rate and increasing spending on health and social protection. It deactivated the clause only after one year; the short duration of the suspension speaks to the credibility of the arrangement, given the severity of the COVID-19 crisis. In contrast, Hurricane Matthew caused widespread damage in 2016 but was deemed not to meet the fiscal threshold and hence did not precipitate suspension of the rule.

The framework corrects specific institutional weaknesses that had led to deficit overshooting in the past. The minister of finance is obliged to submit to Parliament a fiscal responsibility statement describing the overall strategy. The minister is also required to submit a fiscal management strategy that reports and explains deviations between fiscal targets and outcomes over the preceding year and projects the government's finances over the coming three fiscal years, together with a macroeconomic framework outlining the assumptions behind these revenue and spending estimates. The independent auditor general is then tasked with examining the ministry's reports and providing an assessment to Parliament within six weeks of the ministry's submission.[32]

This framework addressed the problem of excessive public sector wage growth by requiring the Ministry of Finance to describe a specific trajectory for bringing public sector wages down to 9 percent of GDP by the end of fiscal year 2016. Together with concurrent amendments to the Public Bodies Management and Accountability Act, it required public bodies to prepare and submit information on their financial performance in the current and preceding years, together with explanations for deviations from budget, to be used as input for the fiscal responsibility statement. The framework

32. An important observation that bears on the question, asked below, of whether lessons from Jamaica generalize is that the auditor general is a strong institution and office, given this constitutional guarantee.

enforces a time limit for these submissions and subjects them to independent assessment by the auditor general. This rigorous and transparent framework allowed the deficits of these public entities to be brought down from an average of 2 percent of GDP in the 2006–2011 period to about 0.5 percent of GDP in 2013–2022.[33]

In sum, the Fiscal Responsibility Framework provided concrete numerical targets for debts and deficits, along with associated deadlines and a well-defined escape clause; required the minister of finance to provide multiyear plans for how the targets will be achieved; mandated the transparency of assumptions and forecasts, together with independent assessments by the auditor general; and held the central government and public bodies accountable for revenue shortfalls and expenditure overruns.

III.B. Institutionalized Partnership and Monitoring

The failure of fiscal adjustment efforts in 2010–2012 indicates that the rules adopted in 2010 by themselves were not enough. There remained a significant danger of the process being derailed until EPOC was created in 2013 and until EPOC was supported by the signing of a meaningful national partnership agreement—the Partnership for Jamaica Agreement (NPC 2013) that same year. The Partnership for Jamaica Agreement affirmed that the government, political opposition, and social partners had reached a consensus on policy priorities; it committed the parties to monitoring the conformity of public policies with those priorities. EPOC meanwhile enabled financial stakeholders to track fiscal policies and hold the government accountable for its budgetary actions. We think of the NPC, which produced the Partnership for Jamaica Agreement, as a consultative and consensus-building institution designed to create confidence that the burden of fiscal adjustment was equitably shared—as an example of the approach to consensus building known in the literature as "democratic corporatism." We think of EPOC primarily as a monitoring and information dissemination technology focused on the budget.[34]

The NPC in fact drafted a series of partnership agreements, some of which were more substantive than others. The first such agreement in 2011

33. See IMF Article IV reports. In addition, the government agreed to privatize CAP as a condition of its programs with the IMF and in 2020 merged its holdings with those of General Alumina Jamaica, which is owned and operated by the Hong Kong-based Noble Group; 55 percent of the merged entity was owned by Noble Group, 45 percent by the government of Jamaica. Jamaica Urban Transit, in contrast, remains government owned and operated (see https://www.jamalco.com/about-us.html).

34. In practice there was overlap between the objectives and deliberations of the two entities, as we make clear below.

was a mere "code of conduct" including no specific commitments.[35] The political opposition consequently boycotted its signing, indicative of a lingering lack of trust. The 2013 Partnership for Jamaica Agreement, which coincided with the inauguration of sustained debt reduction, was very much more detailed. It was the outcome of an extended round of consultations on specific issues, including debt. The document started by acknowledging the sense of crisis created by "inter alia, an unsustainable debt-to-GDP ratio" (NPC 2013, 3). It spoke of the need for social dialogue and participatory decision making to engender "trust and confidence among the Partners" (ibid., 3). It provided commitments by both the government and the opposition to the principles of transparency, accountability, and consultation and to the pursuit of "long-term national goals rather than short-term political imperatives" (ibid., 5); by business to limit profit margins; from trade unions to address problems of low productivity; and by representatives of civil society to help "stabilise and transform the economy" (ibid., 6). It then presented four specific policies requiring monitoring and accountability, of which "Fiscal Consolidation (with Social Protection and Inclusion)" (ibid., 6) had priority of place.

The NPC agreed to monitor the compliance of parties to the terms of this agreement in a manner complementary to the other newly created oversight body, EPOC, which focused more closely on fiscal functions. EPOC was established specifically to reassure domestic holders of sovereign bonds that the government would keep to its fiscal commitments, including the rules set out under the fiscal Responsibility Framework. The government had completed a first domestic debt exchange in 2010, as noted above, as a precondition for the 2010 IMF Stand-by Arrangement. But that arrangement was offtrack already in early 2011, due to an overrun of the 9 percent public sector wage/GDP target. The prime minister resigned in October, and his party was immediately voted out of office, raising questions about its successor's intentions. The new government then tabled a second domestic debt exchange, also described above, with an eye toward securing a new IMF agreement.[36] This time, however, debt holders refused to participate absent assurances that any additional maturity extensions and coupon reductions would be the last. Hence the creation of EPOC to monitor implementation

35. It had simply listed a set of "key guiding principles" such as sensitivity, courage, patience, and understanding.

36. This involved tapping the IMF's Extended Fund Facility (EFF), which provides assistance to countries with medium-term as opposed to short-term balance of payments problems because of structural issues or slow growth. Jamaica's 2013 EFF arrangement was for four years.

of the government's economic reform measures and specifically its compliance with IMF targets and conditions.

EPOC has eleven members representing the public sector, trade unions, business, and finance, with relatively heavy representation of this last category. This difference in composition compared to the NPC—specifically, greater representation of financial interests—reflects EPOC's focus on fiscal questions.[37] EPOC issues reports, typically quarterly, on fiscal policy conduct and outcomes, comparing realized tax revenues and expenditures with those budgeted and analyzing their determinants. It has continued to do so since the country's ongoing arrangement with the IMF concluded in 2019. This is a key observation: monitoring was shared with the IMF virtually from the start, and it has continued long since the IMF exited the scene.

EPOC's assessments are posted on its website, together with communiqués and video recordings by its chair. In addition, EPOC started a program called "On the Corner" that involved going from town to town with reports in hand, explaining what the debt reduction program was designed to achieve. These consensus-building efforts were followed by a visible improvement in public opinion: survey data from the Latin American Public Opinion Project show little change between 2006 and 2014 in the share of the public thinking that the economic situation was improving and then a steady increase after 2014.[38]

Recently, the government and Parliament agreed to provide a proper legal basis and full independence for its proceedings by creating a fiscal commission to "provide an informed second opinion on fiscal developments and . . . play a constructive role in informing the public and, in so doing, incentivizing adherence to Jamaica's fiscal rules" (Clarke 2023, 17). EPOC will stop meeting once the fiscal commission is fully staffed and operational in fiscal year 2024.

III.C. Ownership

Jamaica was under IMF programs in 2010–2011 and earlier, but those programs went offtrack. They did not result in sustained debt reduction.

37. At the same time, EPOC has sufficiently broad nonfinancial sector representation to effectively supplement the dialogue and consensus-building efforts of the NPC. Members engage in dialogue and consensus building that allows the principal stakeholders to monitor and express their views regarding the conformity of fiscal policies with shared public priorities of fiscal accountability and equitable burden sharing.

38. The precise question asked is, "Do you think the country's current economic situation is better than, the same as or worse than it was 12 months ago?" See https://www.vanderbilt.edu/lapop/.

This earlier experience and the experiences of myriad other countries are reminders that IMF involvement is no guarantee of success.

The difference in Jamaica starting in 2013 involves that oft-mentioned but rarely explained, or even defined, concept of ownership. By ownership we mean that country authorities and, importantly, stakeholders to whom those authorities are accountable develop and maintain a broad and credible commitment to the agreed program of policies.[39] In Jamaica, the commitment was broad because it was based on an encompassing partnership agreement that the burden of adjustment would be widely and fairly shared. It was credible because policies and outcomes could be benchmarked against concrete rules and thresholds and because there existed institutionalized monitoring mechanisms to verify the compliance of stakeholders with their commitments.

Well-defined rules and robust partnerships made for ownership of the country's fiscal adjustment and IMF programs. Jamaican officials successfully completed the first ten quarterly reviews under the 2013–2017 Extended Fund Facility (EFF) arrangement. Even when there was a change of government from the PNP to the JLP in March 2016, debt reduction continued. The new JLP administration successfully completed the eleventh, twelfth, and thirteenth quarterly reviews with the IMF and then surprised all concerned by announcing the early ending of the EFF and immediately entering a precautionary Stand-by Arrangement.[40] When the IMF and Jamaican authorities held the High-Level Caribbean Forum in Kingston in November 2017, leaders of both political parties endorsed institutionalizing EPOC. The following April the cabinet embraced the concept of an independent fiscal institution. One month later, the minister of finance delivered a speech, "Enhancing Jamaica's Fiscal Responsibility Framework" (Clarke 2018), initiating another consultative process designed to transfer responsibility for budgetary monitoring from the ad hoc body EPOC to a permanent, independent fiscal commission.

III.D. Origins

The question is how Jamaica was able to reduce political polarization and achieve a broad social consensus in favor of debt reduction. And how

39. Boughton (2003) is one of the rare sources providing an actual definition along these lines.

40. Precautionary arrangements are for cases when countries do not intend to draw on the IMF facility but retain the option of doing so.

and why did it create the institutionalized partnerships that were central to this process?

Here again, our answer has two parts. The first element is Jamaica's historical journey: its troubled history as an independent nation and the lessons drawn from that early experience by political leaders and the public. Over time, that experience and those lessons translated into a visible decline in political polarization and political violence. The second element is the construction of institutions for monitoring, consensus building, and cohesion, first in the electoral realm, where the need was most pressing, but then in the areas of economics, finance, and finally fiscal policy, where policymakers could build on earlier precedents and achievements.

Jamaica was not always a cohesive society. Shortly after independence, Yale sociologist Wendell Bell observed of the country: "The white upper classes, the brown middle classes, and the black lower classes are grossly unequal, with economic and social advantages accruing most to the upper and least to the lower classes" (1964, 38). This sense of inequality fueled the PNP's 1972 electoral victory and its subsequent populist rhetoric and policies. One year before the 1976 election in which PNP Prime Minister Michael Manley won a second term, he declared: "Jamaica has no room for millionaires." For those who wanted to be millionaires, he suggested, "we have five flights a day to Miami" (Levi 1990, 157). In response to the PNP's rhetoric and policies, the opposition JLP moved farther to the right. Accusations of electoral intimidation, malfeasance, and fraud were widespread (Electoral Commission of Jamaica 2014). Political violence soared: election season saw rampant shootings in Kingston's "garrisons" of those thought to favor the political opposition. Estimates are of more than a hundred politically motivated murders in 1976 and more than 800 in 1980.

This ghastly situation created a groundswell for reducing political polarization and violence. Prominent public figures took the lead: during the One Love Peace Concert, before an audience of more than 30,000, the country's leading artist, Bob Marley, joined hands onstage with the prime minister and the leader of the opposition. Following their defeat in the 1980 election, Manley and the PNP moderated their rhetoric and policies. On retaking office in 1989, the PNP embraced the JLP's previously implemented economic reforms, as noted in section II. Manley himself articulated the party's new more collaborative, centrist approach to economic policy:

> [The PNP], like many other people in the broad social democratic movement, placed greater reliance at that time on the capacity of the state to be a direct factor in production. Experience showed us that the state is not necessarily a reliable

> intervener in production. You stretch your managerial capacity and create tensions with the private sector that can be counterproductive. So the second great lesson that we learned is not really to depend on the government as a factor in production but rather to use government as an enabling factor for the private sector. (Massaquoi 1990, 112)

Given Manley's personal popularity, his party's endorsement of this newfound economic policy consensus played an important role in creating a less polarized political environment, more conducive to constructive engagement. This is evident in figure 2, where we see discrete steps down in political polarization after 1980 and again after 1989.

The second element was institution building. To address problems of electoral intimidation and fraud, leaders of both parties agreed that oversight of elections should be removed from the direct ministerial control of the government. Following the recommendations of a bipartisan commission, Parliament voted in 1979 to create an independent, nonpartisan institution with representation of both political parties and civil society to monitor and validate electoral results. This electoral advisory committee (EAC) consisted of eight members: the director of elections, three members of civil society, and four nominated members (two each from the JLP and PNP).[41] The EAC was "not answerable to any minister of government" (Electoral Commission of Jamaica 2014, 21). It was a venue for dialogue between the parties and other stakeholders and had independent authority to invalidate any election result tainted by violence or malfeasance.[42]

The EAC was a first step on Jamaica's journey toward social partnership. It was the precedent for creating, over the next three decades, other independent, multistakeholder consultative bodies that addressed not electoral intimidation and fraud but other issues, notably including economic growth and debt reduction. These subsequent bodies are listed in table 3.

The National Planning Council in 1989 was the next significant institutional innovation: its twenty-two members brought together government officials with business, trade union, and other private sector members (representing academic, professional, and consumer interests) in monthly

41. Civil society representatives were selected by the governor-general. The governor-general, a legacy of the British Commonwealth, represents the monarch on ceremonial occasions and has various powers, sporadically exercised, under the constitution. The EAC was unlike other standing commissions, such as the public service commission and police commission, in that the director-general took advice directly from both the prime minister and the leader of the opposition and not just from the prime minister.

42. For a detailed discussion of the workings of the EAC and the process by which it was created, see Electoral Commission of Jamaica (2014, 20–40).

Table 3. History of Partnership Agreements

1979	**Electoral Advisory Committee (EAC)** Nonpartisan body established to monitor elections, consisting of representatives of the Electoral Office of Jamaica, each of the two major political parties and civil society.
1989	**National Planning Council** Multisector body established to advise government on issues related to national planning.
1997	**ACORN** Social dialogue group led by members of civil society.
2003	**Partnership for Progress** Initiated by the Private Sector Organization of Jamaica
2008	**National Social Partnership Consultative Committee** Creation of National Social Partnership Consultative Committee including representatives of government, parliamentary opposition, private sector, trade unions and civil society groups
2009	**National Partnership Council (NPC)** Creation of National Partnership Council consisting of representatives of the government, parliamentary opposition, and other stakeholder groups. NPC engages in respectful, constructive, and sustained dialogue and collaborates on critical national economic and social issues. Established under the operating rubric of **Partnership for Transformation**, the NPC, has operated across successive administrations. It led further to the creation of the following partnerships.
2011	**Partnership Code of Conduct**
2013	**Partnership for Jamaica**
2016	**Partnership for a Prosperous Jamaica**
2022	**Partnership for Jamaica's Strong and Sustainable Recovery**

Source: Jamaica Office of the Prime Minister (2024).

meetings intended to "contribute to the formulation of economic policies and programmes, to assess economic performance and to identify measures designed to achieve broad-based development and growth in productivity, employment and the national product" (Government of Jamaica 1989, 1).

The National Planning Council was followed in 1997 by ACORN, a venue for social dialogue "in which leaders of the Country's labour unions, private sector and academia have met together continuously over the last twenty-one years, focusing on building social capital and trust among actors in key sectors of the Jamaican society in pursuit of national growth and competitiveness" (Wint 2018). The launch of ACORN again coincided with a visible drop in political violence and a drop in political polarization centered on 1999. ACORN is widely viewed as a progenitor of the partnership committees and councils culminating in creation of the National Partnership Council in 2009, as described in section III.B. Creation of the NPC was followed by one of Jamaica's largest post-independence declines in political polarization and political violence (see figure 2). This became the vehicle

Table 4. Events Surrounding Creation and Operation of the Economic Programme Oversight Committee

2010	Jamaica Debt Exchange (January 14–February 3)
	Stand-by Agreement with IMF begins (February 4)
	Fiscal Responsibility Framework introduced (February 22)
2011	Stand-by Agreement with IMF goes offtrack and is ended
	Prime Minister Bruce Golding of JLP steps down (October)
	Golding is succeeded by Andrew Holness of JLP
2012	PNP wins general election in January
	Debt-to-GDP ratio peaks at 144 percent
2013	Economic Programme Oversight Committee (EPOC) created
	National Debt Exchange (February 12)
	IMF Extended Fund Facility agreement begins (May 1)
2014	Fiscal Responsibility Framework augmented (April 1)
2016	JLP wins election (February), continues with EPOC etc.
	IMF Extended Fund Facility successfully completed (November 10)
	Precautionary Stand-by Agreement with IMF begins (November 11)
2017	IMF managing director hosts high-level IMF Caribbean forum in Kingston
2018	Independent Fiscal Commission Consultative Body announced
2019	Precautionary Stand-by Agreement with IMF completed (no money drawn); Lagarde praises Jamaica's successful conclusion of program across two administrations and reducing debt-to-GDP ratio by 50 percentage points: https://jis.gov.jm/former-imf-boss-praises-jamaica/
2020	COVID-19: Timeline for reducing debt-to-GDP ratio to 60 percent extended from 2026 to 2028
2023	Independent Fiscal Commission established to succeed EPOC (March 7)
	Jamaica's debt rating upgraded by S&P to BB- (September)
	Jamaica issues first international bond in local currency (November)

Source: Authors' compilation.

for the landmark Partnership for Jamaica Agreement in 2013 and its sequel, the Partnership for a Prosperous Jamaica, when the government changed hands in 2016.

Building on this foundation, Jamaican leaders used this same approach of building encompassing institutions with independent powers starting in 2010 when the issue became fiscal adjustment and debt sustainability. Table 4 shows the sequence of institutional steps, starting with introduction of the Fiscal Responsibility Framework in 2010 and continuing with creation of EPOC in 2013. A sense of crisis informed the decision to create EPOC in 2013, just as a sense of crisis informed the decision to create the electoral advisory commission in 1979. In 1979, political violence had threatened Jamaica's survival as a political democracy. In 2013, normalizing the finances was "essentially a matter of survival of the Jamaican nation as a viable nation state," as Peter Phillips, the minister of finance, put it (Wigglesworth 2020, par. 15) .

The generous representation of financial interests in EPOC was important for disciplining and creating confidence in fiscal and financial policies, as argued above. Jamaica's specific approach to debt restructuring had a lot to do with the development of this particular institutional configuration. Governments are typically more inclined to restructure external than domestic debts.[43] Historically, domestic debt has been held by residents, who are also citizens and voters. Incumbent governments prefer to avoid subjecting them to financial pain, knowing that those voters can retaliate by inflicting electoral pain. In addition, where domestic debt is held by local financial institutions, there can be fear that restructuring it could destabilize the financial system. In Jamaica, unusually, a combination of practicalities and legal restrictions made it more expedient to restructure domestic debt. This meant that local financial institutions, which held this debt, became highly attentive to fiscal developments. Because the painful 2010 restructuring was unsuccessful, in that it did not help to put the country on the path to sustained debt reduction, local financial institutions refused to participate in the deeper 2013 restructuring without further reassurance. They viewed the creation of EPOC, their ample representation, and the efficient operation of its monitoring and consultation functions as a nonnegotiable precondition for their participation in this second round.

While EPOC had relatively heavy representation of financial interests and focused on monitoring fiscal policies and outcomes, including those associated with the IMF's EFF, it did not do so to the exclusion of other issues, such as collective bargaining. The unions had agreed to a two-year public sector wage freeze as part of the failed 2010 Stand-by Agreement. Just as investors were now willing to accept further maturity extensions and coupon reductions only as part of a successful program, unions were prepared to extend the wage freeze only if they were confident that the broader stabilization program had a reasonable chance of success. Their representation on EPOC was important for creating this confidence. In the words of Phillips, the monitoring and deliberations of EPOC "did much to build public support across class lines, and I dare say, across political lines for the necessity of the fiscal consolidation and pro-growth efforts at public sector reform and legislative reforms" (Phillips 2017, 2). As further explained by Clarke (2018, 11),

> the consensus building mechanisms of non-governmental bodies had, and continue to have, an indispensable role to play. It was against this background that the previous administration approached members of the financial community with a second debt exchange and the unions with a multi-year wage freeze as prior actions for entry into the Extended Fund Facility. Both groups correctly insisted

43. Though not always: see Reinhart and Rogoff (2011).

> on the right to monitor Jamaica's economic program in return for such sacrifices, in order to ensure that Jamaica maintained its commitments to the reforms embedded in the agreement with the IMF. And so EPOC was born.

This passage makes clear that while the focus of EPOC monitoring was fiscal policy and Jamaica's commitments to the IMF, the committee entailed a broader social partnership in the manner of the other multipartner consultative bodies that preceded it. And while EPOC's establishment coincided with the country's entry into the EFF, the impetus for its creation came from Jamaica. As IMF managing director Christine Lagarde noted in 2014, monitoring of an IMF reform program by an "outside group . . . is something that I have never heard of [and] that none of my staff had heard of" (Wigglesworth 2020, par. 19).

IV. Do the Lessons Generalize?

Does the Jamaican case generalize? Can other economies similarly shed heavy debt burdens by strengthening fiscal rules and backing them with consensus-building institutions? The IMF evidently thinks so: its current managing director has pointed to Jamaica as a model to be followed (Georgieva 2019).[44] At the same time, the fact that Jamaica's case is widely seen as exceptional raises questions about whether the lessons generalize. Insofar as the relevant agreements and institutions were products of Jamaica's distinctive history, shouldn't they be treated as sui generis?

We address these questions through a discussion of two countries, Ireland and Barbados, that bear a strong resemblance to Jamaica in their success at putting in place consensus-building arrangements accompanied by fiscal rules.

IV.A. Ireland

Ireland already had strong fiscal institutions, but these were further strengthened in 1987. The budgetary process was centralized and disciplined. The government first debated the minister of finance's budget proposal in a series of meetings. When the taoiseach (prime minister) exercised strong discipline over his spending ministers, free-riding was contained. To

44. Similarly, her predecessor, Lagarde, in the interview just quoted, went on to suggest that "this is surely a role model that should be emulated elsewhere. With everybody inside the tent, all voices are heard, and everyone has a stake in success" (Wigglesworth 2020, par. 19).

this end, in May 1987 the Fianna Fáil government led by Taoiseach Charles Haughey set up an Expenditure Review Group, a kind of "star chamber" made up of two finance department officials and an independent economist. Finance department staff first drew up a list of schemes that were candidates for termination or funding cuts. The department secretary then was called before the review group, where he was expected to agree to the finance department's list or offer his own proposals for abolishing schemes and saving money. Ministers failing to find the necessary cuts were subject to ruthless discipline by the taoiseach, who threatened them with political consequences.[45]

Under the constitution, only the government could propose spending and tax plans, and there could be no amendments in parliamentary debates; this limited the logrolling characteristic of other legislatures. The government's tax proposals might be voted down by coalition partners or when it was a minority relying on independents. But in 1987 the leader of the opposition agreed not to oppose budgets that promised to address the country's now pressing debt and deficit problems, so adoption of the government's austerity budget was assured.

Despite these institutional arrangements, previous governments' budget-balancing efforts had proved unavailing. Uncoordinated strikes by the country's myriad craft unions first secured substantial pay increases, to which public sector unions then responded with aggressive wage demands of their own (Sexton and O'Connell 1997). Budgets made provision for limited public sector pay increases but were then blown off course by demands for substantial increases from public sector unions, requiring additional expenditure during the year.[46] The 1984 Building on Reality plan had the modest goal of reducing the primary deficit sufficiently to just stabilize the debt at its then high level but was upended in 1986 by a teachers' strike to which the government capitulated. Governments sometimes responded with additional steps to balance the budget, but weak growth undermined the fiscal accounts.

By 1987 a deeply unfavorable interest rate/growth rate differential had contributed to an alarming rise in the public debt ratio to 110 percent of

45. As Haughey put it in a letter to ministers, "any Minister who came to the Cabinet with proposals for expenditure should bring his seal of office with him [i.e., be prepared to resign] and any Department Secretary who proposed expenditure would be sacked" (quoted in Cromien 2011, par. 3).

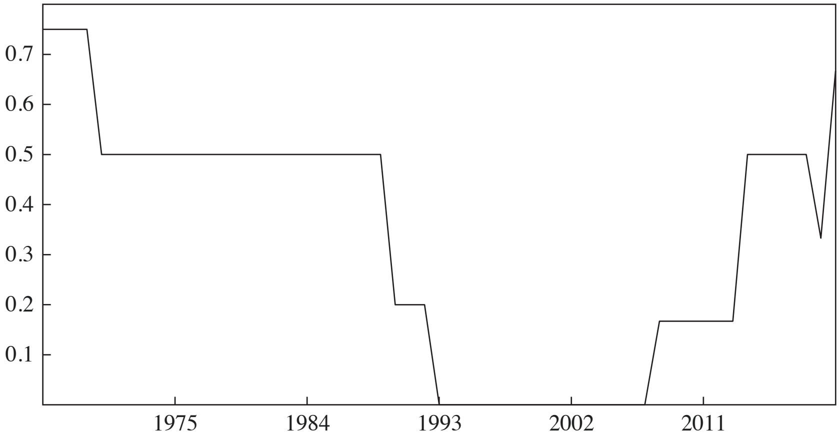
46. As also happened in Jamaica after 2010.

GDP.[47] This led the new Fianna Fáil government to take a different tack, seeking to forge a consensus with trade unions and employers' associations. As the political party historically associated with centralized bargaining, it started by negotiating a common agenda with the unions, whose leaders agreed to pay restraint in return for cuts in taxes on labor income, increased say in decision making, and initiatives to foster job creation. Coordination was facilitated by the fact that all but a few unions were affiliated with the Irish Congress of Trade Unions, their umbrella organization (Hogan 2010). Employers' organizations reluctantly came on board, attracted by the prospect of pay moderation but worried that agreement with public sector unions to reduce the length of the workweek might spread to the private sector. Consultations on the details were conducted with farmers, community representatives, and NGOs. The resulting Programme for National Recovery, covering 1988–1990, entailed agreement to limit annual pay increases to 2.5 percent, reduce taxes on employers and employees, and curtail public sector employment through attrition while preserving the overall value of social welfare benefits and essential public services. It encouraged the belief that the sacrifices required for debt reduction would be widely and equitably shared.

These consensus-building arrangements were buttressed by encompassing discussions, by independent analysis to confirm the accuracy of assumptions, and by mechanisms providing ex post verification that everyone was keeping their word. The National Economic and Social Council (NESC), an independent body whose members included business representatives, union leaders, and academics, was enlisted to analyze the realism of the proposed agreement. A Central Review Committee (CRC) with representation of government and the social partners was established to monitor implementation, enabling the parties to verify that everyone was adhering to the agreement. As MacSharry and White (2000) observe, the regular meetings of the CRC enabled the social partners to have continuing input into government decision making. They allowed union representatives to connect concessions on pay restraint to the provision of public services. And they provided "valuable political and economic education" (MacSharry

47. See Kenny (2016). Figures here for Ireland use gross domestic product to scale debt (for consistency with other countries). The alternative would be to use gross national income, given the importance of profits booked in Ireland by multinational corporations. Another alternative is modified gross national income, which subtracts depreciation of intellectual property and leased aircraft as well as the net factor income of redomiciled publicly listed companies. This however would complicate international comparisons and does not change the narrative.

Figure 10. Ireland: Measure of Political Polarization



Source: V-Dem Database (version 13).
Note: Average of survey responses between zero and 4; lower figure indicates less polarization.

and White 2000, 130). These arrangements were not unlike consultation and consensus-building institutions adopted in Jamaica and were accompanied by a decline in measured political polarization (figure 10).

As in Jamaica, this cooperative burden-sharing agreement did not come out of nowhere. It did not reflect a sudden realization that the country faced a fiscal crisis; the backdrop of fiscal problems was well known. Rather, it built on earlier proposals. In 1982, a national economic plan, The Way Forward, had proposed a collaborative approach to eliminate the budget deficit within four years, but governments were unable to implement it, as described above. In 1986 the NESC then issued a report recommending shared fiscal adjustment, but the unions again refused to participate, and the coalition was again unable to implement it.

What then was different in 1987? First, the Thatcher reforms in the United Kingdom were a wake-up call for the unions, which were forced to recognize the need to balance pay and productivity. With Margaret Thatcher's defeat of the miners' union, confrontation with employers and the government no longer appeared to be a successful way forward. Second, earlier agreements had focused on the need for wage restraint to the exclusion of other factors; incorporating tax and workplace considerations into the 1987 agreement brought labor on board. Third, at this point, finally, "all the parties, through their earlier involvement with the NESC, were familiar with the scale of the problems facing the economy" (MacSharry and White

2000, 129). This answer to the question of why 1987 was different is a reminder that, as in Jamaica, history and experience matter.

Almost immediately, deficits narrowed, and the debt ratio began falling. Real net borrowing by the public authorities fell by half between 1987 and 1988; it again fell by half between 1988 and 1989 (Honohan 1992). The success of the Programme for National Recovery led to a series of subsequent agreements, each covering three years. The government was able to sustain large primary budget surpluses for an extended period. Despite the fact that it took time for growth to pick up and for the interest rate/growth rate differential to become favorable, the public debt ratio fell from its peak of 110 percent in 1987 to barely 60 percent a decade later, and then to a scant 20 percent a decade after that.[48]

Success has many fathers. Other observers will point to rapid catch-up growth, aid from the EU's structural funds, and Ireland's success at attracting foreign investment. While not disagreeing, we would emphasize solid fiscal institutions and consensus-building arrangements.

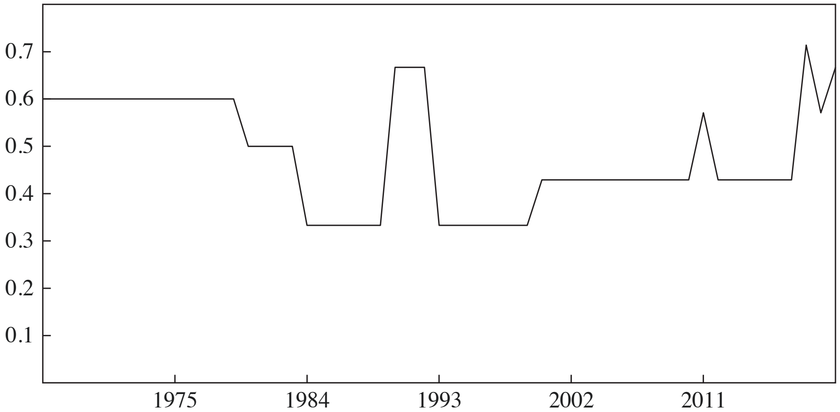
IV.B. Barbados

A last case painting a more mixed picture is Barbados. In July 1991, Prime Minister Erskine Sandiford faced dwindling reserves and a rapidly rising debt-to-GDP ratio. Rather than accepting the IMF's recommendation to devalue the currency, he proposed an 8 percent cut in public sector wages. The Congress of Trade Union and Staff Associations of Barbados responded with a plan exploring other options. However, talks broke down when the prime minister disregarded the congress's proposal and presented public sector workers with a plebiscite that gave them a choice between a wage cut and the IMF-recommended devaluation. Reflecting the national attachment to the currency peg (in operation since 1975) as a nominal anchor—especially given the evidence of the inflation spike following Jamaica's 1991 exchange rate liberalization—workers opted, somewhat remarkably, for the wage cut.

The government implemented these reductions on October 1, 1991 (IMF 2021). On October 24 and again on November 4–5, some 30,000 congress protesters, the proportional equivalent of 36 million Americans, marched through the streets of Bridgetown calling for the prime minister's resignation. The congress challenged the wage cut in court, arguing that the

48. Kenny (2016) shows that $r - g$ contributed negatively to debt reduction until the mid-1990s, after which Ireland's growth accelerated to the high single digits, inaugurating the "miracle" period. Ireland in its earlier years thus resembled Jamaica in that the success of debt reduction did not hinge on rapid growth and a favorable interest rate/growth rate differential.

Figure 11. Barbados: Measure of Political Polarization



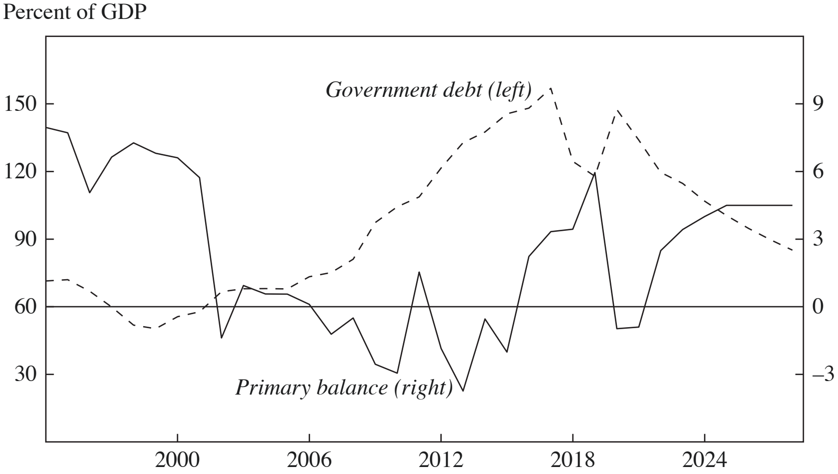
Source: V-Dem Database (version 13).
Note: The largest increase in political polarization occurs around the 1991 public sector wage cut. Average of survey responses between zero and 4; lower figure indicates less polarization.

government had negotiated in bad faith and violated the constitution. The case was escalated to the Privy Council of Barbados.

Simultaneously, Barbados experienced its first post-independence increase in political polarization (figure 11). The deterioration in political conditions was not lost on IMF staff. The minutes of the IMF's July 1992 executive board meeting make clear staff's approval of the government's willingness to cut wages but express concerns about its ability to sustain the wage agreement given societal tensions. The minutes also note staff's strongly held view of the need for the private sector to accept wage restraint for the stabilization plan to succeed.

From the time of the wage cut through the signing of the Stand-by Arrangement, only the government and public sector unions engaged in meaningful discussions; the private sector was notably absent. Meanwhile, the debt-to-GDP ratio continued to rise. Finally, in August 1993, a three-party agreement known as "Protocol for the Implementation of a Prices and Incomes Policy" (Government of Barbados 1995) was brokered with help from the Anglican Church. Employers agreed to limit price increases, accept lower profits, and share their financial accounts with the unions. In return, private sector unions assented to a two-year wage freeze (retroactive to April 1993) and agreed to keep demands for future pay raises in line with increases in productivity. The government committed not to devalue, and all parties agreed to create a national productivity board to provide data on which to

Figure 12. Barbados: Government Debt and Fiscal Balance, 1994–2028



Source: IMF staff projections and the World Economic Outlook Database (October 2023).

Note: In fiscal years, which run from April 1 to March 31. Figures for 2023–2028 are projections from the 2023 Article IV report (December 2023). Government debt on a net basis.

base future negotiations (Henry and Miller 2009). There followed a marked decline in political polarization between 1993 and 1994 (figure 11). Barbados ran a primary budget surplus of 8 percent of GDP in 1994 and a primary surplus in excess of 5 percent of GDP in each of the next five years. As a result, the net debt-to-GDP ratio came down from 71 percent in 1994 to 50 percent by 1999 (figure 12).

Beyond that, however, the process did not last. At the turn of the century the debt ratio began rising again, rapidly with the onset of the global financial crisis when growth stagnated and the interest rate/growth rate differential turned especially unfavorable. The debt-to-GDP ratio rose from 61 percent of GDP in 2000 to as high as 157 percent in 2017.

Part of the problem was that the consensus-building measures of the mid-1990s were not buttressed by significant reforms of fiscal institutions (increased fiscal transparency, independent institutions for monitoring the realism of budgeting assumptions, explicit fiscal rules). The government continued to make unbudgeted transfers to loss-making state-owned enterprises (SOEs) providing water, transportation, electricity, waste disposal, and health services. These transfers averaged 7.5 percent of GDP per annum in the decade following the 2008 global financial crisis (IMF 2021). They culminated in an IMF program and debt restructuring in 2018.

At this point, Barbados finally put in place an explicit debt-to-GDP target and measures enhancing the transparency and facilitating outside monitoring of the fiscal accounts, including the operations of SOEs. The Financial Management and Audit Act was amended to give expenditure ceilings to line ministries. The amendment enhanced monitoring and supervision of SOEs by adding internal audit and reporting requirements. The government committed to a target for its debt of 60 percent of GDP and a path for the primary balance consistent with getting there by 2034 (delayed for two years by COVID-19-related financial disruptions). These fiscal rules complemented and reinforced the existing social partnership agreement.

Barbados appears to be emulating the Jamaican model by forming a committee, with the participation of private sector business associations and labor unions, to monitor implementation of its 2018 Barbados Economic Reform and Transformation Plan and by establishing an independent fiscal council to monitor and advise on fiscal policy implementation.[49]

A difference between Barbados and Jamaica is that Barbados undertook a comprehensive debt restructuring in 2018–2019 that entailed significant present-value reductions and encompassed external as well as internal debt. A new government initiated the restructuring in its first week in office, immediately ahead of a large external payment and leveraging its ability to blame its predecessor for the need for exceptional measures. The authorities were anxious to reach a loan agreement with the IMF, and the IMF, not allowed to lend to a government with an unsustainable debt, required the restructuring as a condition.

Barbados had the advantage that its global bonds contained collective action clauses (unlike Jamaica's some years earlier), the global campaign to encourage their inclusion having gained traction over time. Compared to Jamaica, its external debt thus could be restructured more quickly, given less scope for free-riding and litigation. Domestic debt was far and away the most important component of the government's obligations, however, and domestic debt securities did not include collective action clauses. But because the bonds were governed by domestic law, these provisions could be retrofitted by an act of Parliament.

The resulting net present value loss for the creditors was as much as 44 percent on external debt and 43 percent on domestic debt (Anthony, Impavido, and van Selm 2020). Recall how in Jamaica there had been a

49. This makes Barbados and Jamaica the only two Caribbean countries with independent fiscal councils. Like its Jamaican counterpart, the BERT Monitoring Committee continues to issue regular public reports.

reluctance to impose restructuring-related losses on the banks for fear of causing financial instability. In Barbados, more than 40 percent of domestic debt was again held by the banks.[50] But all five Barbadian banks were foreign owned.[51] All five banks were strongly capitalized, had healthy parents, and could absorb losses. Again, the message—which emanates also from Jamaica's contrasting experiences in the 1990s and after 2009—is that a sound financial system is important for successful debt reduction.

Ireland and Barbados, like Jamaica, are small economies, consistent with the idea that consensus building is easier where there is a limited number of agents. They are sectorally specialized, open economies highly exposed to exogenous shocks, consistent with the argument that achieving this kind of adjustment-facilitating consensus is especially urgent in a shock-prone environment. Ireland is more ethnically and socioeconomically homogeneous than Jamaica, consistent with the literature suggesting that a neo-corporatist approach to consensus building is easier when cooperation is not complicated by ethnic divisions (Katzenstein 1985; Gavrilets, Auerbach, and van Vugt 2016). Jamaica, as a society with more income and wealth inequality, and more racially and ethnically diverse historical roots, had to work for decades to construct an economic and social consensus in favor of debt reduction.

It is not clear that large countries can easily follow the small country strategy of partnership and engagement to reduce political polarization and build consensus. But neither is it clear that they will be able to reduce their debts without it.

V. Conclusion

There is no questioning the desirability of bringing down high public debt-to-GDP ratios. Heavy debts prevent governments from increasing expenditure and cutting taxes in recessions and emergencies (Romer and Romer 2019). Debt-service burdens limit the scope for productive public spending (Jalles and Medas 2022). Especially when they are short in term or denominated in foreign currency, large debts are a source of financial fragility.

Given the magnitude of inherited debts, meaningful debt reduction can be achieved only by running substantial budget surpluses for extended periods. At present, $r - g$ differentials have turned less favorable, given

50. Excluded from this calculation is debt held by the public sector itself (principally the National Insurance Scheme and the central bank).

51. Three big ones were owned by AAA-rated Canadian financial institutions, the two smaller ones by banks headquartered in oil-rich Trinidad and Tobago.

upward pressure on real interest rates—reflecting investors' higher required return to hold additional government securities—and the troubled outlook for global growth.[52] Debt restructuring, never a panacea, has grown more fraught and complex with the substitution of market finance for official finance and the emergence of nontraditional creditors.[53]

Yet only a small handful of countries have succeeded in running the requisite large primary surpluses for extended periods. Jamaica, having cut its debt-to-GDP ratio from 144 percent of GDP in 2012 to 72 percent in 2023, is a prime case in point. This makes it important to understand the Jamaica exception.

Meaningful debt reduction was accomplished only when Jamaica put in place two prerequisites: (1) a set of rules anchoring fiscal policy, which allowed investors and others to monitor government policies and assess their conformance with projections; and (2) a partnership agreement creating confidence that the burden of adjustment would be widely and fairly shared. Both elements were needed. Jamaica had experimented previously with partnership agreements, but these alone did not prevent debt from exploding. Jamaica adopted fiscal rules three years before the start of its debt reduction process, but these rules did not prevent debt from continuing to rise.[54] Together, however, the two elements launched Jamaica on a debt reduction course whose success few countries have been able to match.

The lessons from Jamaica's experience with fiscal rules, we suggest, generalize to other countries. Jamaican officials adopted simple numerical targets for the debt-to-GDP ratio, with dates attached. The finance minister was tasked with formulating a multiyear budget detailing how the debt ratio would get from here to there. Parliament strengthened the governance of state-owned enterprises and public bodies to avoid cost overruns. The fiscal rules included a state-of-the-art escape clause that balanced flexibility with credibility. And an auditor general whose independence was constitutionally guaranteed provided outside verification of the government's claims. These lessons can be adopted elsewhere.

The other element of the recipe, encompassing partnership agreements, is more difficult to replicate. EPOC and the Partnership for Jamaica Agreement

52. Kose and Ohnsorge (2024) forecast a further slowdown in trend growth in emerging markets and developing economies over the next five years. There is of course no agreement on how much growth will slow and real interest rates will rise. These issues are discussed in Arslanalp and Eichengreen (2023).

53. The failure of more than a small handful of governments to reach restructuring agreements under the G20's Common Framework for Debt Treatments illustrates the point.

54. And even before that the country had been subject to IMF-negotiated fiscal targets.

that launched and kept Jamaica on the path of debt reduction were products of a distinctive national learning process that began a third of a century earlier with the Electoral Advisory Commission, whose structures and processes were transferred to other domains, including, eventually, the budgetary. The decision to start down this road reflected the country's history of race and class division and political violence, away from which leaders and society turned at the end of the 1970s when the country reached the political brink. Other heavily indebted countries have different political histories. They do not all face the same dire political circumstances. Nor is there any guarantee that their leaders and publics will respond in the same way.

Our analysis and the literature on democratic corporatism suggest that encompassing partnership agreements such as Jamaica's are most prevalent in smaller countries, where it is easiest to get the stakeholders around a table. They are most prevalent in small, open, sectorally specialized economies where vulnerability to external shocks is high and cooperation on adjustment is urgent. They are most prevalent where interest group negotiations are relatively structured and centralized. They are easiest to reach in relatively homogeneous societies not riven by class or racial divisions.

These observations leave us relatively pessimistic about the efficacy of fiscal rules in countries such as Germany, whose provisions lack flexibility. They leave us skeptical about the enforceability of the EU's revised fiscal rules, which lack simplicity to accompany flexibility, and where their imposition from outside raises questions about ownership and enforceability (Eyraud and others 2018). And they leave us concerned about the scope for sustained debt reduction in large countries like the United States with high levels of political polarization.

At the same time, Jamaica's experience suggests that societal divisions are endogenous. They can be modified over time, not least through the creation and operation of encompassing institutional partnerships. And these partnerships can be deployed to create fiscal rules with the simplicity, flexibility, and acceptance needed to be enforceable and effective.

What it takes to modify societal divisions and to usefully deploy, during crises, the increase in social capital that flows from a more cohesive society brings us to the final lesson from Jamaica's experience: the importance of leadership. Our discussion of the earlier period emphasized the critical role of Prime Minister Manley's intellectual shift in favor of economic and fiscal pragmatism. In terms of more recent experience, one could similarly point to the strong leadership of Finance Minister Peter Phillips before 2016 and Finance Minister Nigel Clarke thereafter. Economists prefer to ground their arguments in institutions and market forces rather than personalities. But such institutions presuppose leaders with the vision and character to use them for the good of the country. Without leadership, there is no broad acceptance to accompany credibility and solidify ownership. The World Bank's Growth Commission (Brady and Spence 2010) identified leadership as one of the five common traits of countries with sustained high growth in the post-World War II period. The same might be said of public debt reduction for small and large countries alike.

ACKNOWLEDGMENTS We thank Laura Alfaro, Eleanor Brown, Patrick Honohan, Tracy Robinson, Jón Steinsson, Emil Verner, DeLisle Worrell, and Karina Garcia, Thomas Pihl Gade, Vitor Gaspar, Jan Kees Martijn, Pablo Morra, Constant Lonkeng Ngouana, Uma Ramakrishnan, Bert van Selm, and Esteban Vesperoni (all IMF) for helpful discussions.

References

Alesina, Alberto, Roberto Perotti, and José Tavares. 1998. “The Political Economy of Fiscal Adjustments.” *Brookings Papers on Economic Activity*, no. 1: 197–248.

Alesina, Alberto, and Guido Tabellini. 1990. “A Positive Theory of Fiscal Deficits and Government Debt.” *Review of Economic Studies* 57, no. 3: 403–14.

Anthony, Myrvin, Gregorio Impavido, and Bert van Selm. 2020. “Barbados’ 2018–19 Sovereign Debt Restructuring—A Sea Change?” Working Paper 2020/034. Washington: International Monetary Fund.

Arslanalp, Serkan, and Barry Eichengreen. 2023. “Living with High Public Debt.” *Economic Policy Symposium Proceedings: Structural Shifts in the Global Economy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Arslanalp, Serkan, and Takahiro Tsuda. 2014. “Tracking Global Demand for Emerging Market Sovereign Debt.” Working Paper 2014/039. Washington: International Monetary Fund.

Bell, Wendell. 1964. *Jamaican Leaders: Political Attitudes in a New Nation*. Berkeley: University of California Press.

Boughton, James M. 2003. “Who’s in Charge? Ownership and Conditionality in IMF-Supported Programs.” Working Paper 2003/191. Washington: International Monetary Fund.

Brady, David, and Michael Spence. 2010. *Leadership and Growth: Commission on Growth and Development.* Washington: World Bank Group.

Clarke, Nigel. 2018. “Enhancing Jamaica’s Fiscal Responsibility Framework.” Ministry of Finance and the Public Service. https://www.mof.gov.jm/wp-content/uploads/hmfps-speech-fiscal-responsibility-framework.pdf.

Clarke, Nigel. 2023. “Preserving, Increasing and Sharing the Gains of Economic Recovery: Opening Budget Presentation 2023.” Ministry of Finance and the Public Service. https://www.mof.gov.jm/wp-content/uploads/HMFPS-BUDGET-PRESENTATION_March-7-_TO-PRINT-E.pdf.

Cromien, Seán. 2011. “The Story of the First ‘Bórd Snip.’” *Irish Journal of Public Policy* 3, no. 2: 99–106.

Electoral Commission of Jamaica. 2014. *History of the Electoral Commission of Jamaica*. Kingston: Electoral Commission of Jamaica.

Eyraud, Luc, Debrun, Xavier, Andrew Hodge, Victor Duarte Lledó, and Catherine A. Pattillo. 2018. “Second-Generation Fiscal Rules: Balancing Simplicity, Flexibility, and Enforceability.” Staff Discussion Note 2018/004. Washington: International Monetary Fund.

Gavrilets, Sergey, Jeremy Auerbach, and Mark van Vugt. 2016. “Convergence to Consensus in Heterogeneous Groups and the Emergence of Informal Leadership.” *Scientific Reports* 6: 29704.

Georgieva, Kristalina. 2019. “The IMF’s Purpose: Delivering for People.” Plenary address to the International Monetary Fund Annual Meeting, Washington, October 18. https://www.imf.org/en/News/Articles/2019/10/18/sp101819-md-am-plenary-speech.

Government of Barbados. 1995. "Protocol for the Implementation of a Prices and Incomes Policy: 1995–1997." https://labour.gov.bb/pdf/Protocol%202.pdf.

Government of Jamaica. 1989. "The National Planning Council." Ministry Paper 18/89. Kingston: Government of Jamaica.

Government of Jamaica. 2013. "Fiscal Policy Paper FY 2013/14." Kingston: Government of Jamaica.

Henry, Peter Blair. 2013. *Turnaround: Third World Lessons for First World Growth.* New York: Basic Books.

Henry, Peter Blair. 2023. "Productive Public Investment: An Innovative Framework for Aligning Jamaica's Economic and Social Priorities." Lecture delivered at the Dialogue for Development Series, Planning Institute of Jamaica, Kingston, November 2. https://www.pioj.gov.jm/product/productive-public-investment-an-innovative-framework-for-aligning-jamaicas-economic-and-social-priorities/.

Henry, Peter Blair, and Conrad Miller. 2009. "Institutions versus Policies: A Tale of Two Islands." *American Economic Review* 99, no. 2: 261–67.

Hogan, John. 2010. "Economic Crises and the Changing Influence of the Irish Congress of Trade Unions on Public Policy." In *Irish Business and Society: Governing, Participating and Transforming in the 21st Century*, edited by John Hagan, Paul F. Donnelly, and Brendan K. O'Rourke. Dublin: Gill Books.

Honohan, Patrick. 1992. "Fiscal Adjustment in Ireland in the 1980s." *Economic and Social Review* 23, no. 3: 285–314.

International Monetary Fund (IMF). 2000. *Jamaica: Staff Report for the 1999 Article IV Consultation*. Country Report 2000/008. Washington: International Monetary Fund.

International Monetary Fund (IMF). 2018. *Jamaica: 2018 Article IV Consultation, Third Review under the Stand-By Arrangement and Request for Modification of Performance Criteria*. Country Report 2018/103. Washington: International Monetary Fund.

International Monetary Fund (IMF). 2021. *Barbados: 2021 Article IV Consultation, Sixth Review under the Extended Arrangement under the Extended Fund Facility, and Request for Modification Performance Criteria*. Country Report 2021/268. Washington: International Monetary Fund.

International Monetary Fund (IMF). 2023. *Jamaica: 2022 Article IV Consultation—Press Release; Staff Report and Staff Statement*. Country Report 2023/083. Washington: International Monetary Fund.

Jalles, João Tovar, and Paulo A. Medas. 2022. "Economic Growth after Debt Surges." Working Paper 2022/159. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2022/07/29/Economic-Growth-After-Debt-Surges-521357.

Jamaica House of Parliament. 2010. *An Act Further to Amend the Financial Administration and Audit Act.* Kingston: Jamaica House of Parliament. https://www.japarliament.gov.jm/attachments/341_The%20Financial%20Administration%20and%20Audit%20(Amendment)%20(No.2)%20Act,%202010.pdf.

Jamaica House of Parliament. 2014. *An Act to Amend the Financial Administration and Audit Act.* Kingston: Jamaica House of Parliament. https://www.japarliament.gov.jm/attachments/341_The%20Financial%20Administration%20and%20Audit%20Act%202014.pdf.

Jamaica Office of the Prime Minister. 2024. "National Partnership Council." https://opm.gov.jm/national-partnership-council/.

Kandil, Magda, Qiaoe Chen, Xin Li, Genevieve Lindow, Mario Mansilla, Joel Okwuokei, Marika Santoro, Jochen Schmittmann, and Solomon Stavis. 2014. "Labor Market Issues in the Caribbean: Scope to Mobilize Employment Growth." Working Paper 14/115. Washington: International Monetary Fund.

Katzenstein, Peter J. 1985. *Small States in World Markets: Industrial Policy in Europe.* Ithaca, N.Y.: Cornell University Press.

Kenny, Sean. 2016. "Public Debt Episodes in Irish Economic History 1950–2015." Lund Paper in Economic History: General Issues, no. 2016/145. Lund: Department of Economic History, Lund University. https://lucris.lub.lu.se/ws/portalfiles/portal/8159815/Episodes_in_Irish_Public_Debt_1950_2015_LUP.pdf.

King, Damien. 2001. "The Evolution of Structural Adjustment and Stabilisation Policy in Jamaica." *Social and Economic Studies* 50, no. 1: 1–53.

Kirkpatrick, Colin, and David Tennant. 2002. "Responding to Financial Crisis: The Case of Jamaica." *World Development* 30, no. 11: 1933–50.

Kirton, Claremont, and James Ferguson. 1992. *Jamaica: Debt and Poverty.* London: Oxfam.

Kose, M. Ayhan, and Franziska Ohnsorge, eds. 2024. *Falling Long-Term Growth Prospects: Trends, Expectations and Policies.* Washington: World Bank.

Laeven, Luc, and Fabian Valencia. 2020. "Systemic Banking Crises Database II." *IMF Economic Review* 68: 307–61.

Langrin, R. Brian. 2013. "Policy Lessons from Postmortems of Jamaica's Two Recent Debt Exchanges." Working Paper. Kingston: Bank of Jamaica. https://boj.org.jm/uploads/pdf/papers_pamphlets/papers_pamphlets_Policy_Lessons_from_Postmortems_of_Jamaica%27s_Two_Recent_Debt_Exchanges.pdf.

Leon, Hyginus, and Rasheeda Smith. 2012. "Fiscal Management for Sustainable Development." International Monetary Fund. https://boj.org.jm/uploads/pdf/papers_pamphlets/papers_pamphlets_Fiscal_Management_for_Sustainable_Development.pdf.

Levi, Darrell E. 1990. *Michael Manley: The Making of a Leader.* Athens: University of Georgia Press.

MacSharry, Ray, and Padraic White. 2000. *The Making of the Celtic Tiger: The Inside Story of Ireland's Boom Economy.* Dublin: Mercier Press.

Massaquoi, Hans. 1990. "Interview with Jamaican Prime Minister Michael Manley." *Ebony* 45: 110–18.

Ministry of Finance and the Public Service, Government of Jamaica. 2023. "Jamaica Launches First Ever Jamaican Dollar $-Linked Instrument in International Capital Markets." Press release, November 7. https://jis.gov.jm/jamaica-launches-first-ever-jamaican-dollar-linked-instrument-in-international-capital-markets/.

National Partnership Council (NPC). 2013. “Partnership for Jamaica Agreement.” July 31. https://jis.gov.jm/media/PFJ-Agreement-final2_2013July31.pdf.

Okwuokei, Joel Chiedu, and Bert van Selm. 2017. “Debt Restructuring in the Caribbean—The Recent Experience.” In *Unleashing Growth and Strengthening Resilience in the Caribbean,* edited by Trevor Serge Coleridge Alleyne, Inci Ötker, Uma Ramakrishnan, and Krishna Srinivasan. Washington: International Monetary Fund.

Phillips, Peter. 2017. “Forward.” In *Communiqués of the Economic Programme Oversight Committee*. Kingston: EPOC.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2011. “The Forgotten History of Domestic Debt.” *Economic Journal* 121, no. 552: 319–50.

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.

Schmid, Juan Pedro. 2016. “Addressing Debt Overhang: Experiences from Two Debt Operations in Jamaica.” Policy Brief 259. Washington: Inter-American Development Bank.

Sexton, J. J., and P. J. O’Connell, eds. 1997. *Labour Market Studies: Ireland.* Luxembourg: Office for Official Publications of the European Communities.

Stephens, Evelyne Huber, and John D. Stephens. 1986. *Democratic Socialism in Jamaica: The Political Movement and Social Transformation in Dependent Capitalism.* Princeton, N.J.: Princeton University Press.

Stone, Carl, and Stanislaw Wellisz. 1993. “Jamaica.” In *The Political Economy of Poverty, Equity, and Growth: Five Small Open Economies*, edited by Ronald Findlay and Stanislaw Wellisz. New York: Oxford University Press.

Valencia, Oscar, Carolina Ulloa-Suárez, and Jorge Guerra. 2024. “Navigating Fiscal Flexibility: Assessing the Quality of Escape Clauses in Rule-Based Fiscal Frameworks.” Forthcoming. Washington: Inter-American Development Bank.

Wigglesworth, Robin. 2020. “Inside the IMF’s Outrageous, Improbably Successful Jamaican Programme (Pt. 2).” *Financial Times*, February 16. https://www.ft.com/content/e1f0cf23-6a83-4c61-a880-d7984c5a8f6f.

Wint, Alvin. 2018. “Lest We Forget—Jamaica’s Path to Economic Stability.” *Jamaica Gleaner*, December 6. https://jamaica-gleaner.com/article/news/20181207/alvin-wint-lest-we-forget-jamaicas-path-economic-stability.

Comments and Discussion

COMMENT BY

LAURA ALFARO Arslanalp, Eichengreen, and Henry advance two empirical regularities: "Sharp, sustained reductions in public debt are exceptional, especially recently" and "only rarely have [governments] succeeded in bringing those higher debt ratios back down once the emergency passed."

The paper then analyzes Jamaica's fascinating case to draw lessons in managing sustained primary fiscal surpluses. As Reinhart, Reinhart, and Rogoff (2015) describe, sustained debt reduction can involve: orthodox measures—economic growth (relation $g - r$), primary surplus (fiscal adjustment/austerity); and heterodox measures—surprise inflation tax (implicit default of a sudden surprise burst in inflation if debt in local currency is not indexed), explicit default or restructuring, and financial repression (also for domestic-issued debt). As the authors note, Jamaica reduced its debt-to-GDP burden from 144 percent to 72 percent with modestly favorable $r - g$ and more than its fair share of external shocks. Jamaica stands out because it succeeded the "old-fashioned way" through primary surpluses. As the authors show, only a handful have succeeded via fiscal surplus. This rarity makes understanding these exceptions essential.

The authors then answer the *how* and *why* of the sustained debt reduction beyond crisis, that is, even after the emergency had passed. The case of Jamaica highlights two features: (1) fiscal rules that are transparent and clear (with numerical debt, fiscal balance, and public sector wages targets) and flexible budgetary rules within the Fiscal Responsibility Framework introduced in 2010 and augmented in 2014 with monitoring, reporting, and independent verification; and (2) ownership that debt reduction was anchored in Jamaica's tradition of consensus building and social partnerships, which allowed for a sense of transparency and fairness in burden sharing, dialogue, ownership, and continuity. Despite changes in government parties, Jamaica could sustain the benefits beyond the crisis.

The paper then generalizes the lessons with additional cases that have succeeded via fiscal rules and institution-based consensus building also anchored in earlier historical experience, such as Ireland (1980s) and Barbados (1990s).[1] Figure 1 plots the countries listed in table 1 in the paper—Emerging Markets and Developing Economies (EMDEs) with the largest five-year debt reductions—against a polarization measure (level corresponding to the initial year of each country's largest five-year debt reduction). Additionally, figure 2 illustrates the relationship between fiscal surplus (corresponding to the initial year of each country's largest five-year debt reduction) and the polarization measure. The vertical lines highlighting distribution quartiles underscore the rarity of Jamaica's success and also that of Ireland and Barbados.[2]

The paper starts with historical background, then discusses what Jamaica did and how, addressing fiscal rules, institutionalized partnership and monitoring, ownership, and origins (history and institution building). It ends with the question of whether the lessons can be generalized and compares Jamaica's case with that of Ireland and Barbados.

This is an excellent and comprehensive paper, rich in details and footnotes. The paper uncovers the role of the intricacies of societal norms, political legacies, shared mental models, and formal rules that shape political, economic, and social interactions (North 1990), which are crucial for understanding fiscal policy over time and the complexity of debt management. These systems' internal logic, consistency, and timing are highly complex. Congratulations to the authors for this outstanding work. My comments will center around fiscal rules, the key role of ownership, and reduced polarization. I will end with thoughts on the implications for international financial architecture in the current world of high debt, particularly among poorer countries.

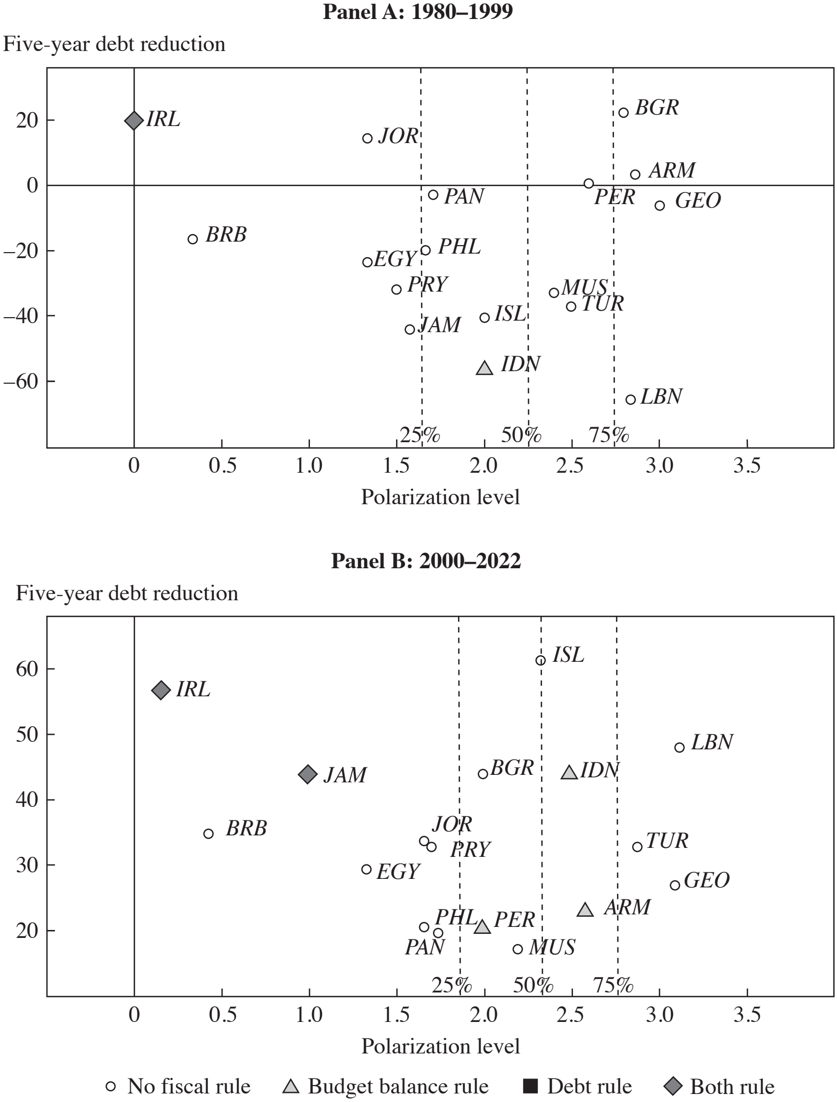
CONTROLLING THE GOVERNMENT: FISCAL RULES "In framing a government which is to be administered by men over men, the great difficulty lies in this: you must first enable the government to control the governed; and in the next place oblige it to control itself."[3]

1. Iceland, after the global financial crisis, is another case.

2. Additionally, markers provide insights into fiscal rules: a diamond indicates countries with both budget balance and debt rule, while a triangle corresponds to the budget balance rule.

3. Bill of Rights Institute, "Federalist 51," par. 1, https://billofrightsinstitute.org/primary-sources/federalist-no-51.

Figure 1. EMDEs with the Largest Five-Year Debt Reductions: 20 Percent Threshold and Polarization



Source: Author's calculations using the IMF general government gross debt (per GDP) to calculate the debt reductions and the IMF government primary balance data for the fiscal surplus; polarization (v2cacamps_mean) is from the V-Dem data set; indicators of budget balance rule and debt rules are from the IMF Fiscal Rule data set from 1985 to 2021.

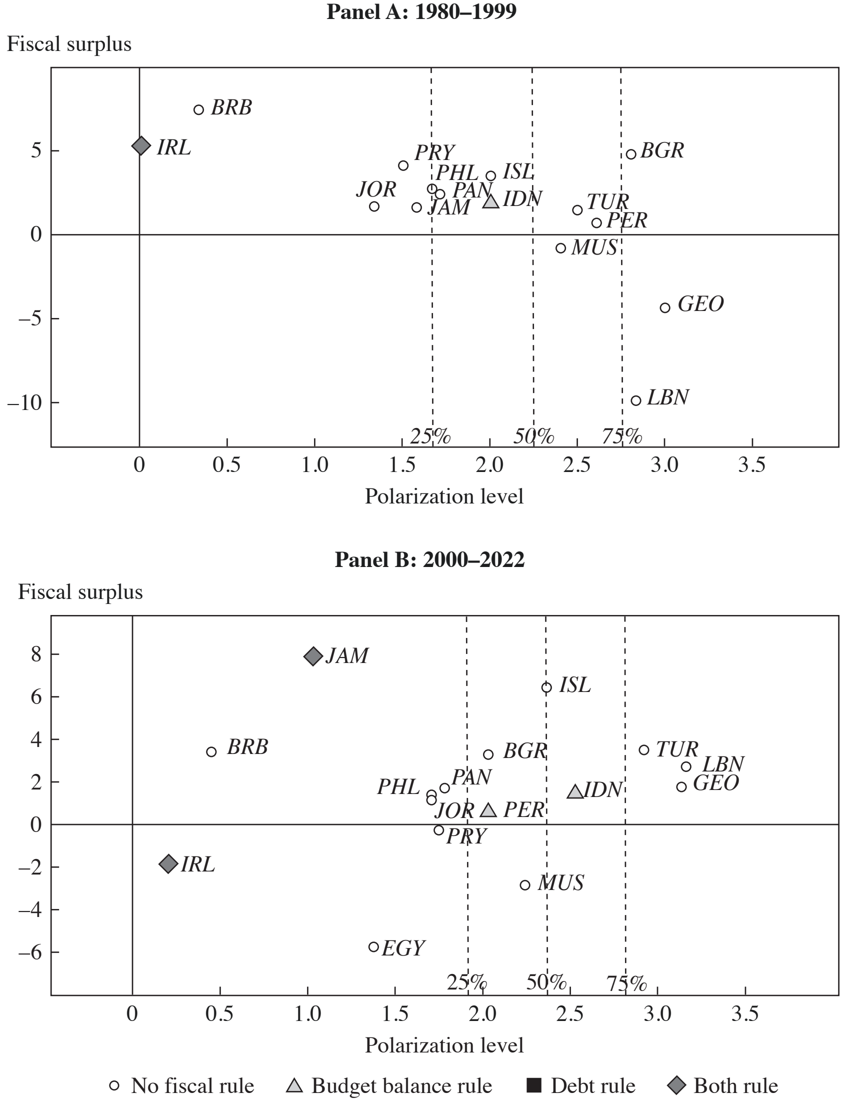
Note: The list of countries is in table 1 of the authors' paper for EMDEs, excluding episodes of external debt restructuring and major oil exporters. The plot includes Ireland, Iceland, and Barbados as well. The plot shows the 25th, 50th, and 75th percentiles of the corresponding V-Dem indicator for all EMDEs as defined by the IMF. https://www.imf.org/en/Publications/WEO/weo-database/2023/April/groups-and-aggregates.

Table 1. Countries in the Data Set (IMF, World Bank, and V-Dem)

ABW	AFG	AGO	ALB	AND	ARE	ARG	ARM	ATG	AUS
AUT	AZE	BDI	BEL	BEN	BFA	BGD	BGR	BHR	BHS
BIH	BLR	BLZ	BOL	BRA	BRB	BRN	BTN	BWA	CAF
CAN	CHE	CHL	CHN	CIV	CMR	COD	COG	COL	COM
CPV	CRI	CUB	CYP	CZE	DDR	DEU	DJI	DMA	DNK
DOM	DZA	ECU	EGY	ERI	ESP	EST	ETH	FIN	FJI
FRA	FSM	GAB	GBR	GEO	GHA	GIN	GMB	GNB	GNQ
GRC	GRD	GTM	GUY	HKG	HND	HRV	HTI	HUN	IDN
IND	IRL	IRN	IRQ	ISL	ISR	ITA	JAM	JOR	JPN
KAZ	KEN	KGZ	KHM	KIR	KNA	KOR	KWT	LAO	LBN
LBR	LBY	LCA	LKA	LSO	LTU	LUX	LVA	MAC	MAR
MDA	MDG	MDV	MEX	MHL	MKD	MLI	MLT	MMR	MNE
MNG	MOZ	MRT	MUS	MWI	MYS	NAM	NER	NGA	NIC
NLD	NOR	NPL	NRU	NZL	OMN	PAK	PAN	PER	PHL
PLW	PNG	POL	PRI	PRK	PRT	PRY	PSE	PSG	QAT
ROU	RUS	RWA	SAU	SDN	SEN	SGP	SLB	SLE	SLV
SML	SMR	SOM	SRB	SSD	STP	SUR	SVK	SVN	SWE
SWZ	SYC	SYR	TCD	TGO	THA	TJK	TKM	TLS	TON
TTO	TUN	TUR	TUV	TWN	TZA	UGA	UKR	URY	USA
UVK	UZB	VCT	VEN	VNM	VUT	WBG	WSM	XKX	YEM
ZAF	ZMB	ZWE	ZZB						

Note: The table lists all the countries included in the data set for figure 3.

Figure 2. EMDEs with the Largest Five-Year Fiscal Surplus and Polarization



Source: Author's calculations using the IMF general government gross debt (per GDP) to calculate the debt reductions and the IMF government primary balance data for the fiscal surplus; polarization (v2cacamps_mean) is from the V-Dem data set; indicators of budget balance rule and debt rules are from the IMF Fiscal Rule data set from 1985 to 2021.

Note: Each point corresponds to the first year of the largest five-year government debt reductions. The countries plotted are those in table 1 of the authors' paper and Ireland, Iceland, and Barbados. The plot shows the 25th, 50th, and 75th percentiles of the corresponding V-Dem indicator for all EMDEs as defined by the IMF. https://www.imf.org/en/Publications/WEO/weo-database/2023/April/groups-and-aggregates.

Lowering high public debt-to-GDP ratios offers significant benefits associated with the costs of fiscal dominance, debt overhang, and crowding out. The literature has also documented the negative effects of costly default (Alfaro and Kanczuk 2005; Mendoza and Yue 2012; Reinhart and Rogoff 2009).

Standard economic theory holds that fiscal policy should be countercyclical (Barro 1979). Yet most emerging countries, possibly owing to distorted political incentives (Alesina, Campante, and Tabelini 2008), follow procyclical fiscal policies, which tend to exacerbate already pronounced cycles (Kaminsky, Reinhart, and Végh 2004). There are many political economy motivations for excessive indebtedness, heterogeneity, wars of attrition over the distribution of costs, common pool problem externalities that lead to a deficit bias, and interest groups.[4] The question relates to the broader rules versus discretion debate on whether a commitment should be required (Halac and Yared 2014).

One solution for fiscal problems is the adoption of fiscal rules. Governments may adopt fiscal rules that constrain their behavior to correct distorted incentives to overspend, particularly in good times. This, in turn, would alleviate distress on rainy days. A data set compiled by the Fiscal Affairs Department of the International Monetary Fund (IMF) identifies countries' adoption of fiscal policy restrictions.[5] Only a handful of countries had fiscal rules in place in 1990; twenty years later, more than 100, which includes different types (debt, deficit, revenues, expenditures).

Do fiscal rules improve welfare? In Alfaro and Kanczuk (2019), we examine the welfare implications of fiscal rules in the context of emerging markets' sovereign debt and default. We transform the traditional sovereign debt and default model by assuming governments' preferences are time-inconsistent and correspond to the quasi-hyperbolic consumption model (Laibson 1997). The consequent conflict between today's government and tomorrow's generates an incentive to precommit to a particular fiscal rule.[6]

We calibrate it to the Brazilian economy, a typical emerging economy. Although large and not an island, three features make Brazil particularly interesting. First, President Dilma Rousseff's impeachment in 2016 was

4. See Alesina and Drazen (1991) and Alesina and Passalacqua (2016) for a literature review.

5. See https://www.imf.org/external/datamapper/fiscalrules/matrix/matrix.htm.

6. Jackson and Yariv (2014, 2015) propose that aggregating citizens' time-consistent preferences naturally results in time-inconsistent preferences that display an extra discount parameter that captures the ex post present bias.

due to disobedience of the existing fiscal rule; second, Congress passed additional fiscal restrictions in December 2016 (Bornhorst and Curristine 2017); and third, the rule was eliminated during COVID-19.

The model can reproduce the Brazilian debt level and default frequency even if the household impatience parameter is calibrated to local interest rates. Some findings include the observation that adopting the optimal fiscal rule implies substantive welfare gains relative to the absence of a rule. Moreover, simpler debt rules can also improve welfare as alternatives to more complex optimal rules. However, not all rules improve welfare; for instance, overly restrictive deficit rules may not reduce welfare. As the Jamaica case highlights, building contingencies into the fiscal rule may be associated with higher welfare. One further point highlighted by the Jamaica case is that increasing transparency and ownership, which can be part of the process of designing and monitoring fiscal rules, indeed helps.[7]

Do fiscal rules control the government? As the paper highlights, not all commitments are effective. Fiscal rules and targets do not always achieve their intended results. The paper mentions the case of the European Union's Stability and Growth Pact. As mentioned, Brazil got rid of the rule. Costa Rica, a small economy (but not an island), recently passed a rule, but basically, policymakers have found ways around it (IMF 2023).

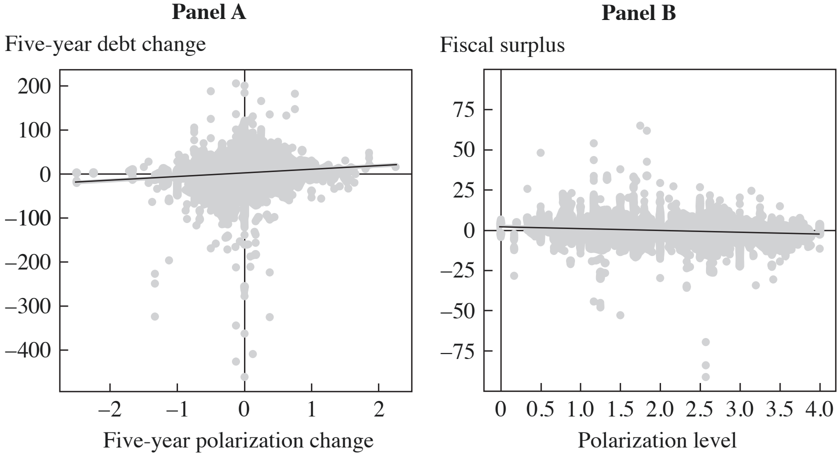
CONTROLLING THE GOVERNMENT: THE SPIRIT OF A PEOPLE "The spirit of a people, its cultural level, its social structure, the deeds its policy may prepare—all this and more is written in its fiscal history" (Schumpeter 1918, 2).

In the case of Jamaica, as noted by the authors, the fiscal rules were adopted already before the start of its debt reduction process, yet they did not prevent debt from continuing to increase. In contrast, the literature tends to see fiscal rules as a way to deal with polarization. As the paper shows, the process is complex and reinforcing, building on the country's history. Factors such as ownership, transparency, reduced polarization, and monitoring of fiscal rules have been crucial in Jamaica's success. In contrast, Brazil's case lacked real buy-in despite the votes.

A strand of the literature studies the effects of polarization and the government's incentives to tax and spend (and how they affect future governments' ability to tax and spend). The literature looks at different forms of polarization, heterogeneity, and conflict of interest (Eslava 2011):

7. The work by Rogoff (1990) and Rogoff and Sibert (1988) shows that fiscal cycles and excess spending can be the outcome of imperfectly informed voters.

Figure 3. All Countries in the Data Set (1980–2022)



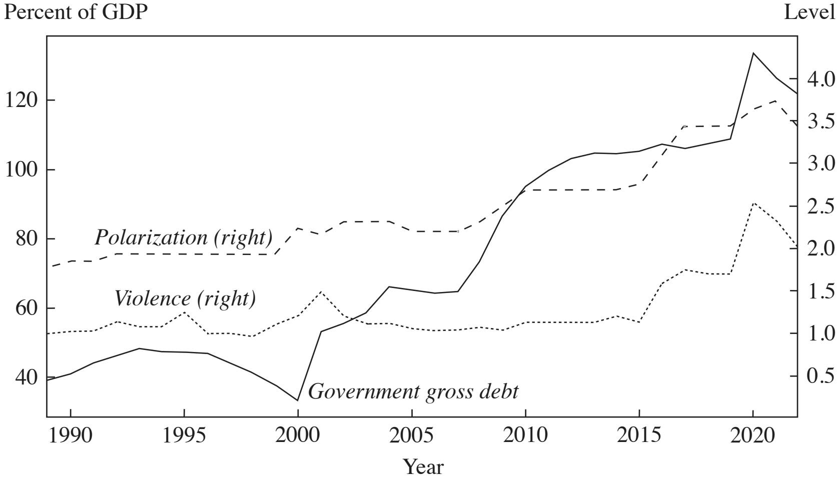
Source: Author's calculations using the IMF general government gross debt (per GDP) to calculate the debt change, the central government debt (per GDP) data from the World Development Indicators (World Bank) to input missing values, and the IMF government primary balance for fiscal surplus; missing fiscal surplus values are filled with net lending (+) / net borrowing (–) (per GDP) of the World Development Indicators (World Bank); polarization (v2cacamps_mean) is from the V-Dem data set.

Note: All countries from IMF and World Bank data sets are utilized for five-year debt change (debt increment, yearly sampled). The polarization change represents the five-year polarization increment prior to the corresponding five-year debt change.

heterogeneity between policymakers and voters, heterogeneity of fiscal preferences across politicians, and heterogeneity of fiscal preferences across social groups or regions. The theoretical work, however, has mixed implications. Although much of the work implies that higher polarization leads to higher deficits, it depends on the assumptions and type of heterogeneity. In Alesina and Tabellini (1990), polarization leads to overspending and deficits; in Persson and Svensson (1989), it depends on the type of incumbent; while Alesina, Baqir, and Easterly (1999) and Azzimonti (2011) state that polarization and disagreement can lead to smaller government and less spending.

As the paper shows, the processes are complex, nonlinear, and unfold over many years. Figure 3, panel A, illustrates the relationship between five-year debt change (debt increment) and polarization change (five-year polarization increment before the corresponding five-year debt change), covering yearly sampled data points from 1980 to 2022. A positive slope suggests that countries experiencing greater decreases in polarization are more likely to reduce their debt levels in the following five years. Similarly,

Figure 4. Polarization in the United States and Debt (1989–2022)



Source: Author's calculations using the IMF general government gross debt (as a percentage of GDP) to calculate the debt change; polarization (v2cacamps_mean) and violence (v2caviol_osp) data are sourced from the V-Dem data set.

Note: The time series plot illustrates changes in the US government gross debt and democracy variables over time. The left *y*-axis indicates the percentage of GDP for government gross debt, while the right *y*-axis shows the levels of polarization and violence for each year.

figure 3, panel B, explores the fiscal surplus and polarization level, providing insights into the tendency of less polarized countries to have higher fiscal surpluses. But again, without an in-depth analysis of the country, one may miss the idiosyncrasies and reinforcing forces.

INTERNATIONAL FINANCIAL ARCHITECTURE: SUSTAINED DEBT REDUCTION BEYOND CRISIS From reading the paper, I walked away somewhat pessimistically, not only of the outlook for the United States (figure 4) but also for other countries, as developing countries face increased debt payments in the coming years in a more complex geopolitical environment. In previous work by Arslanalp and Henry (2006), to HIPC (heavily indebted poor countries), the effects of these interventions to reduce debt were not always encouraging. The authors' arguments have potential implications for how the international financial architecture should handle distressed countries. If relevant institutional steps are almost exclusively taken where there is a sense of crisis, as the Jamaica case highlights, perhaps the IMF's softer stance may affect internal dynamics and create better long-term outcomes. In conclusion, as always with these authors, this paper is a must-read!

REFERENCES FOR THE ALFARO COMMENT

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114, no. 4: 1243–84.

Alesina, Alberto, Filipe R. Campante, and Guido Tabellini. 2008. "Why Is Fiscal Policy Often Procyclical?" *Journal of the European Economic Association* 6, no. 5: 1006–36.

Alesina, Alberto, and Allan Drazen. 1991. "Why Are Stabilizations Delayed?" *American Economic Review* 81, no. 5: 1170–88.

Alesina, Alberto, and Andrea Passalacqua. 2016. "The Political Economy of Government Debt." In *Handbook of Macroeconomics, Volume 2B*, edited by John Taylor and Harold Uhlig. Amsterdam: North-Holland.

Alesina, Alberto, and Guido Tabellini. 1990. "A Positive Theory of Fiscal Deficits and Government Debt." *Review of Economic Studies* 57, no. 3: 403–14.

Alfaro, Laura, and Fabio Kanczuk. 2005. "Sovereign Debt as a Contingent Claim: A Quantitative Approach." *Journal of International Economics* 65, no. 2: 297–314.

Alfaro, Laura, and Fabio Kanczuk. 2019. "Fiscal Rules and Sovereign Default." Working Paper 23370. Cambridge, Mass.: National Bureau of Economic Research.

Arslanalp, Serkan, and Peter Blair Henry. 2006. "Policy Watch: Debt Relief." *Journal of Economic Perspectives* 20, no. 1: 207–20.

Azzimonti, Marina. 2011. "Barriers to Investment in Polarized Societies." *American Economic Review* 101, no. 5: 2182–204.

Barro, Robert J. 1979. "On the Determination of the Public Debt." *Journal of Political Economy* 87, no. 5: 940–71.

Bornhorst, Fabian, and Teresa Curristine. 2017. "Brazil: Making the Spending Rule Work." IMF Public Financial Management (PFM) Blog, October 4. https://blog-pfm.imf.org/en/pfmblog/2017/10/by-fabian-bornhorst-and-teresa-curristine1-brazil-is-emerging-from-a-deep-recess.

Eslava, Marcela. 2011. "The Political Economy of Fiscal Deficits: A Survey." *Journal of Economic Surveys* 25, no. 4: 645–73.

Halac, Marina, and Pierre Yared. 2014. "Fiscal Rules and Discretion under Persistent Shocks." *Econometrica* 82, no. 5: 1557–614.

International Monetary Fund (IMF). 2023. *Costa Rica: Technical Assistance Report—Upgrading the Rule-Based Fiscal Framework.* Country Report No. 2023/097. Washington: Author.

Jackson, Matthew O., and Leeat Yariv. 2014. "Present Bias and Collective Dynamic Choice in the Lab." *American Economic Review* 104, no. 12: 4184–204.

Jackson, Matthew O., and Leeat Yariv. 2015. "Collective Dynamic Choice: The Necessity of Time Inconsistency." *American Economic Journal: Microeconomics* 7, no. 4: 150–78.

Kaminsky, Graciela L., Carmen M. Reinhart, and Carlos A. Végh. 2004. "When It Rains, It Pours: Procyclical Capital Flows and Macroeconomic Policies." *NBER Macroeconomics Annual* 19: 11–53.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112, no. 2: 443–77.

Mendoza, Enrique G., and Vivian Z. Yue. 2012. "A General Equilibrium Model of Sovereign Default and Business Cycles." *Quarterly Journal of Economics* 127, no. 2: 889–946.

North, Douglas C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Persson, Torsten, and Lars E. O. Svensson. 1989. "Why a Stubborn Conservative Would Run a Deficit: Policy with Time-Inconsistent Preferences." *Quarterly Journal of Economics* 104, no. 2: 325–45.

Reinhart, Carmen M., Vincent Reinhart, and Kenneth Rogoff. 2015. "Dealing with Debt." *Journal of International Economics* 96, suppl. 1: S43–S55.

Reinhart, Carmen M., and Kenneth Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton, N.J.: Princeton University Press.

Rogoff, Kenneth. 1990. "Equilibrium Political Budget Cycles." *American Economic Review* 80, no. 1: 21–36.

Rogoff, Kenneth, and Anne Sibert. 1988. "Elections and Macroeconomic Policy Cycles." *Review of Economic Studies* 55, no. 1: 1–16.

Schumpeter, Joseph Aloisius. 1918. "Die Krise der Steuerstaates" ["The Crisis of the Tax State"]. *Zeiqragen azu dem Gebiet der Soziologie*, no. 4. English translation: Wolfgang F. Stolper and Richard A. Musgrave, *International Economic Papers*, no. 4 (1954). https://edisciplinas.usp.br/pluginfile.php/5322348/mod_resource/content/1/Crise%20do%20Estado%20Fiscal.pdf.

COMMENT BY

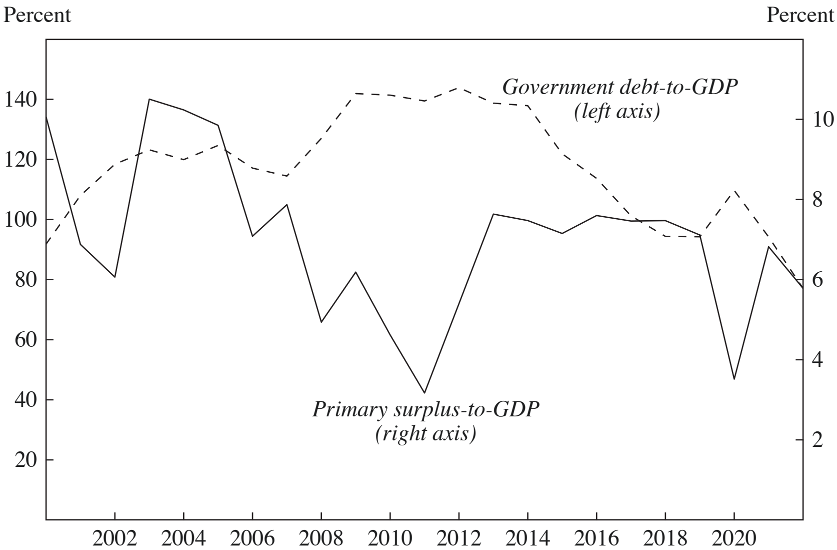
EMIL VERNER The surge in public debt-to-GDP around the world has renewed interest in understanding how countries can bring down public debt ratios. Arslanalp, Eichengreen, and Henry bring our attention to the exceptional case of Jamaica, which slashed its public debt-to-GDP ratio from 144 percent in 2012 to 72 percent in 2023. This case study is especially interesting in the current age of slowing growth and liberalized financial markets, as Jamaica did not "grow its way" out of debt or employ financial repression policies that depressed real interest rates. Instead, debt reduction was achieved the hard way, by running large primary surpluses.

Jamaica's primary surpluses are indeed exceptional. From 2013 to 2019, the primary surplus averaged 7.4 percent of GDP, among the highest in the world.[1] By comparison, Greece's primary surplus averaged 2.2 percent of GDP over the same period, while Germany's was 2.1 percent.[2] Jamaica's feat is even more impressive when one considers that its real GDP only

1. Among countries with a population greater than 1 million inhabitants, Jamaica's primary surplus-to-GDP ratio was only surpassed by Qatar's during this period.

2. Figures are from the International Monetary Fund (IMF) World Economic Outlook Database.

Figure 1. Jamaica's Government Debt-to-GDP and Primary Surplus, 2000–2022



Source: IMF World Economic Outlook Database.

grew by 1 percent per year over the same period. Growth did not serve to reduce the denominator of debt-to-GDP, nor did it facilitate running large primary surpluses.

So how did Jamaica do it? In their valuable account of this fascinating case, Arslanalp, Eichengreen, and Henry provide compelling narrative evidence that Jamaica's primary surpluses were achieved through two factors. The first was fiscal rules that were credible and ambitious, but not overly rigid. The second was consensus building through partnership agreements, which fostered the belief that the burden of adjustment would be fairly distributed in society.

My comments will focus on three points. First, I examine the mechanics of how Jamaica managed to reduce its public debt ratio. Second, I compare Jamaica's experience with other large, sustained debt reductions. Third, I discuss the impact on economic growth. I conclude with some thoughts on the lessons from Jamaica's experience.

UNDERSTANDING THE MECHANICS OF JAMAICA'S DEBT REDUCTION Figure 1 plots Jamaica's government debt-to-GDP ratio and primary surplus-to-GDP ratio. The decline in the debt-to-GDP ratio begins in 2013. The immediate backdrop to the debt reduction was a substantial fiscal consolidation. The primary surplus-to-GDP ratio increased by over 4 percentage points between

2011 and 2013. This was driven by a combination of tax increases and reductions in government spending, especially on public sector wages. These drastic fiscal reforms took place during a period when Jamaica teetered on the edge of a fiscal and financial crisis after Jamaica's earlier International Monetary Fund (IMF) agreement broke down (Johnston 2015). In this respect, I would place more emphasis on the role of fiscal consolidation than do Arslanalp, Eichengreen, and Henry.

Looking at the long-run evolution of the primary surplus-to-GDP ratio in figure 1, one immediately notices that primary surpluses were already large in the decade before the reduction in the debt-to-GDP ratio. Yet these large primary surpluses did not serve to lower the government debt ratio. To understand why, I extend the debt dynamics decomposition in the paper back in time using the following equation:

$$\Delta b_t = d_t + \frac{r_t - g_t}{1+g_t} b_{t-1} + \frac{z_t a_t}{(1+g_t)(1+p_t^*)} b_{t-1} + \frac{(p_t - p_t^*) a_t}{(1+g_t)(1+p_t)(1+p_t^*)} b_{t-1} + residual_t, \tag{1}$$

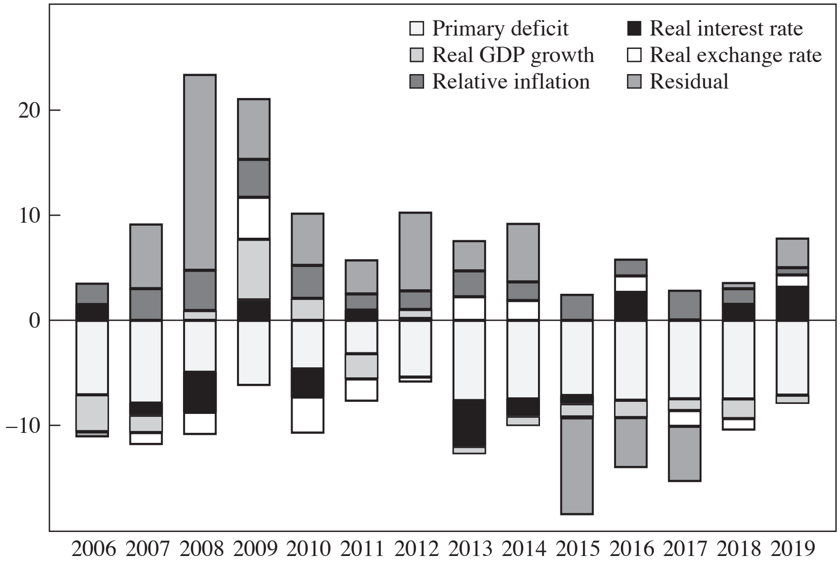
where b_t is the stock of debt, d_t is the primary deficit, r_t is the real interest rate, g_t is the real growth rate, z_t is the real exchange rate depreciation, a_t is the share of foreign currency denominated debt, p_t is the growth rate of the GDP deflator, p_t^* is the growth rate of the US GDP deflator, and $residual_t$ is the stock-flow adjustment.

Figure 2 summarizes the debt decomposition from 2006 to 2019. The debt decomposition for the post-2013 period is very similar to figure 7 in the paper. Looking at the period before 2013, the figure reveals that several factors counteracted the high primary surpluses in the 2000s. The $r - g$ differential contributed to a rise in debt-to-GDP in 2009 through both lower growth (g) and a higher real interest rate (r). The global financial crisis hit Jamaica hard, as exports of natural resources fell sharply (IMF 2010). Interest costs also rose. Furthermore, the share of foreign currency denominated public debt exceeded 30 percent in this period. As a result, real exchange rate depreciation boosted the debt-to-GDP ratio in some years, especially during the global financial crisis.

However, figure 2 reveals that large positive values of the stock-flow adjustment (the residual) was the most important factor for understanding why public debt-to-GDP remained elevated in the 2000s. Moreover,

Figure 2. Dynamic Debt Decomposition—Pesky Residuals

Contribution to Δb in percent of GDP



Source: IMF World Economic Outlook Database.
Note: This figure implements the dynamic debt decomposition in equation (1).

reductions in the stock-flow adjustment played an important role in reducing the debt ratio.[3]

This raises the question: what is captured in the stock-flow adjustment? The stock-flow adjustment can reflect accounting differences across how primary surpluses and public debt are measured. However, it also reflects extra-budgetary expenditures, the realization of contingent liabilities, and losses on state-owned enterprises (SOEs). According to the IMF, repeated fiscal slippages, including losses at SOEs, were at the root of Jamaica's fiscal problems in the 2000s (IMF 2010, 2014). Improvements in the governance and divestment from SOEs mattered for reducing the residual, allowing Jamaica's large primary surpluses to reduce debt (Johnston 2015). This also implies that the extent of fiscal consolidation in the 2010–2013 period was larger than what would be inferred by looking at the primary surpluses alone.

The large role of the stock-flow adjustment is not unique to Jamaica. Campos, Jaimovich, and Panizza (2006) find that this term plays an important

3. The debt sustainability analysis in IMF (2014) also finds large positive residuals contributing to the debt-to-GDP ratio in the 2000s.

role in explaining debt changes in emerging economies and low-income countries, often because of the realization of contingent liabilities. The Jamaica case thus offers an example of a debt reduction where large primary surpluses were not the full story. Rather, the elimination of "below-the-line" contributors to public debt was also important.

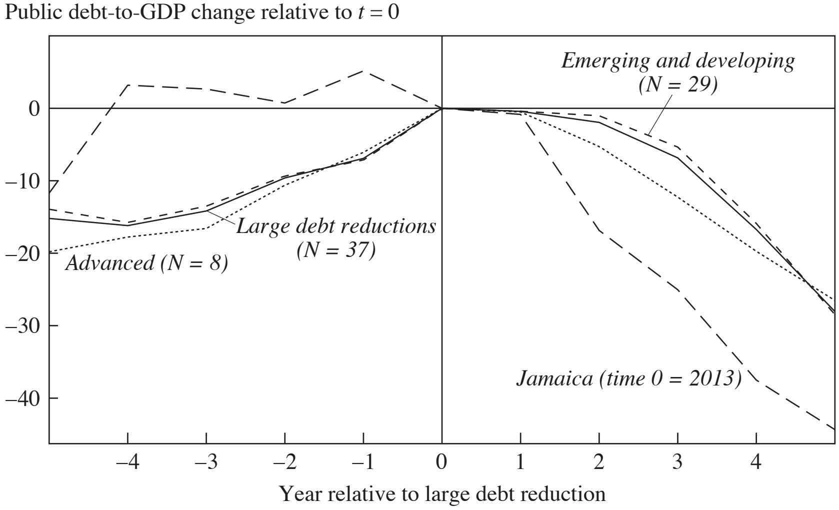
JAMAICA'S DEBT REDUCTION IN AN INTERNATIONAL PERSPECTIVE How does the Jamaica case compare to other large, sustained debt reductions? Is the Jamaican case unique in recent history, or are there parallels? To understand these questions, I turn to an international panel database to identify episodes of major sustained debt reductions. I use the IMF World Economic Outlook Database, which contains an unbalanced panel with information on public debt starting in 1980 and covering up to 193 countries. I define a sustained debt reduction as an episode where the debt-to-GDP ratio declines by at least 20 percentage points over five years. I exclude debt reductions that are accompanied by a default and haircut following the classification in Cruces and Trebesch (2013) and Asonuma and Trebesch (2016). This exercise follows in the spirit of prior work studying large debt reduction episodes (Reinhart, Rogoff, and Savastano 2003; Nickel, Rother, and Zimmermann 2010; Villafuerte and others 2010).

The approach results in a sample of thirty-seven large, sustained debt reduction episodes. The criterion of a 20 percentage point reduction is thus quite stringent. Of the thirty-seven episodes, eight are in advanced economies and twenty-nine are in emerging markets. I also separately analyze advanced economies and emerging markets to understand whether the dynamics of debt reductions differ across the two groups of countries.

Figure 3 plots an event study of the government debt-to-GDP ratio across these large, sustained debt reductions. The level of debt-to-GDP is normalized to zero in event time $t = 0$, the year that the debt reduction begins. The figure also plots Jamaica's debt ratio, with 2013 normalized to event time $t = 0$. The average reduction in public debt-to-GDP amounts to 28 percentage points over five years in this sample of large debt reductions. Jamaica's debt reduction of 44 percentage points over five years is even larger.

Figure 4 plots the average evolution of real GDP (g), the real interest rate (r), the real interest rate minus growth differential ($r - g$), and the primary surplus-to-GDP ratio for the large debt reduction episodes. The typical debt reduction features relatively strong real GDP growth (panel A). This is in sharp contrast with Jamaica, where growth was low throughout its debt reduction. The real interest rate is also negative during the typical debt reduction episode (panel B). This is driven by emerging markets, where high inflation often reduces ex post real interest rates. In Jamaica, as well as

Figure 3. Public Debt-to-GDP Change during Episodes of Large Public Debt Reductions



Source: IMF World Economic Outlook Database.

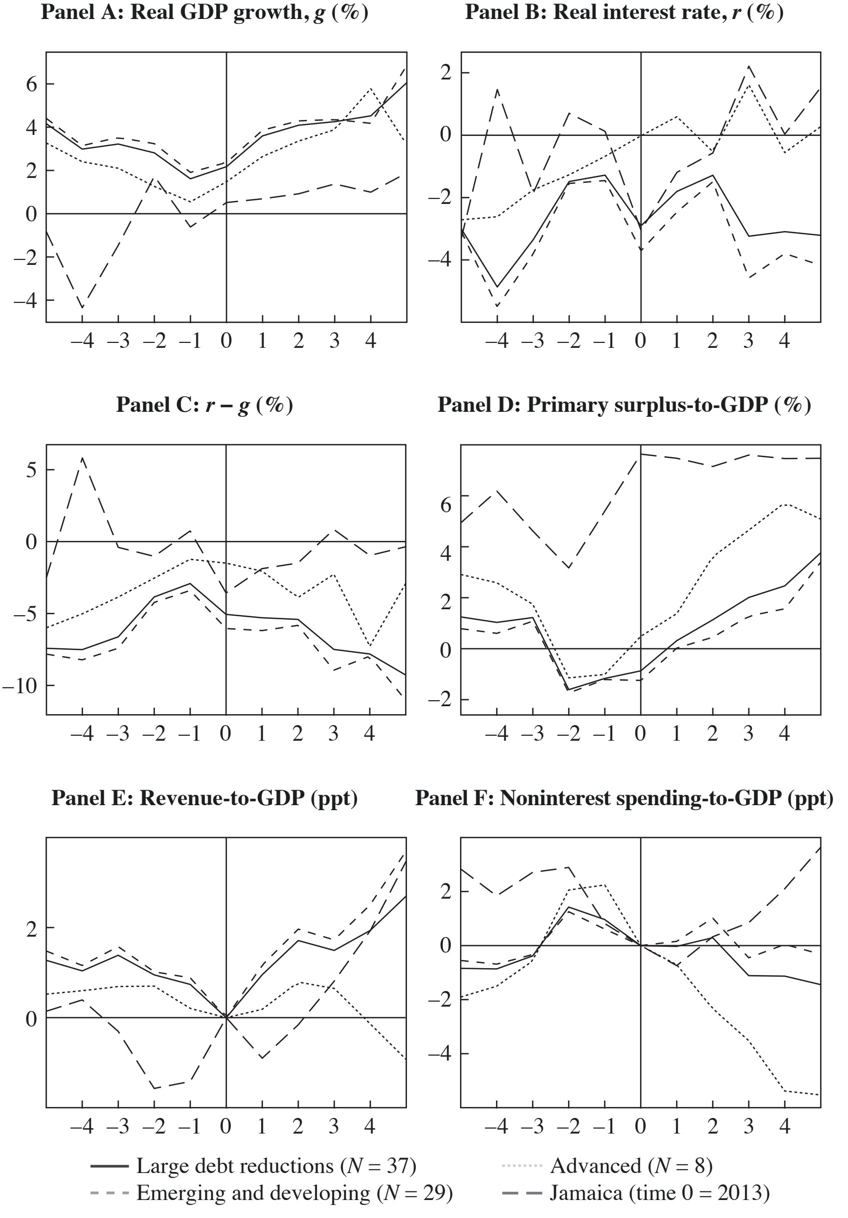
Note: This figure plots the average evolution of public debt-to-GDP across major debt reduction episodes. Public debt-to-GDP is measured as the percentage point change relative to year zero. Time zero is the start of a large public debt reduction, defined as a reduction of at least 20 percentage points over five years. Advanced economies and emerging and developing economies are defined based on the World Economic Outlook country composition.

in advanced economy debt reductions, the real interest rate is not especially low. As a result of strong growth and low real interest rates, the typical debt reduction features highly negative values of $r - g$, which is an important driver of debt reduction (panel C). In contrast, in Jamaica's case the $r - g$ differential was close to zero and thus played a negligible role in its debt reduction.

Panel D in figure 4 shows that the average debt reduction episode involves a rising primary surplus that goes from negative to positive. This is especially the case in advanced economies, where the primary surplus gradually rises by over 6 percentage points relative to GDP.[4] In this sense, Jamaica's debt reduction looks more like the typical advanced economy debt reduction, with two differences. First, the adjustment in the primary surplus was more sudden in Jamaica. Second, the level of the primary surpluses was

4. The outsized role of $r - g$ dynamics in many emerging market episodes of debt reductions is consistent with earlier evidence from Villafuerte and others (2010), who find that in emerging markets $r - g$ often plays as large or even a larger role than the primary surplus. In contrast, in advanced economies, primary surpluses are more important for reducing debt.

Figure 4. Event Study of Large Public Debt-to-GDP Reductions



Source: IMF World Economic Outlook Database.

Note: This figure plots the average evolution of the indicated variable across major debt reduction episodes. Revenue-to-GDP and noninterest spending-to-GDP are the percentage point change relative to year zero, so these two variables are mechanically normalized to zero in event time zero.

much higher. Jamaica's large primary surpluses are especially striking considering its low growth. Running large primary surpluses is easier when strong growth passively raises tax revenues.[5]

Figure 4 also plots the average change in government revenues and noninterest spending as a share of GDP. These values are measured as the percentage point change relative to event time $t = 0$. The path of revenues and spending differs considerably across advanced economies and emerging markets. In advanced economies, the increase in the primary surplus is driven by a reduction in noninterest spending, while revenues are flat or even declining slightly. In contrast, for emerging market episodes the opposite happens: revenues rise as a share of GDP, while spending is relatively flat. In the case of Jamaica, revenues-to-GDP rise, as in other emerging market cases. At the same time, spending was brought down in the run-up to the start of the debt reduction but then began rising again three years into the debt reduction. This did not reduce the primary surplus, as tax revenues also continued to rise.

Prior studies focusing on advanced economies have found that most successful fiscal consolidations are driven by expenditure cuts (Alesina and Perotti 1995; Nickel, Rother, and Zimmermann 2010). However, Jamaica illustrates how fiscal consolidation based partly on an increase in tax revenues can work, especially if revenues are starting from a relatively low initial level, as is often the case in emerging markets and developing countries. At the same time, while revenue increases were important, Jamaica also sharply reduced public sector wages, in line with other successful episodes of debt reduction (Nickel, Rother, and Zimmermann 2010; IMF 2014).

THE IMPACT ON GROWTH The poor growth performance in Jamaica raises the question of whether the extremely tight fiscal policy depressed growth. Research from the IMF finds that fiscal consolidations are often unsuccessful in reducing debt ratios because they harm growth too much (Ando and others 2023). If fiscal consolidation leads to a fall in growth, then debt reduction becomes even more difficult for three reasons. First, lower growth reduces the denominator of the debt-to-GDP ratio. Second, a recession makes it more difficult to run primary surpluses. And, third, it is difficult to maintain political support for fiscal reforms when they cause economic hardship. In the case of Jamaica, a slowdown in growth was highlighted as a key downside risk by the IMF (2010).

On the one hand, growth was very low in Jamaica during the debt reduction, as seen in panel A of figure 4. On the other hand, growth in Jamaica had

5. See, for example, Villafuerte and others (2010).

been low for decades. As the authors note, a structural analysis is required to understand the impact of the tight fiscal policy on growth. However, while we do not know the counterfactual, a casual analysis suggests that the impact of the fiscal contraction in 2011–2013 on growth seems to have been negative but, perhaps, modest relative to the size of the increase in the primary surplus. For example, real GDP growth averaged about 1.8 percent in the 2000–2007 period, but only about 1 percent in the 2012–2019 period according to the IMF World Economic Outlook Database.

It seems likely that tight fiscal policy would have depressed growth by depressing domestic demand. Moreover, the sharp reduction in infrastructure investment could have negative long-term growth consequences (Johnston 2015). However, the negative effects from reductions in demand and potential growth could have been offset by some of the benefits of putting public finances on a more sustainable path, such as improved expectations and reduced private sector borrowing costs (Giavazzi and Pagano 1990). This seems plausible for the case of Jamaica, which was on the verge of a crisis in 2012 when its debt reduction program started. Moreover, fiscal consolidation may have a smaller negative effect on growth in small and highly open economies such as Jamaica, where much of the reduction in demand leaks abroad (Farhi and Werning 2016). Understanding the impact of sustained tight fiscal policy on growth is an important question—both for the case of Jamaica and more broadly.

Despite sluggish growth, an interesting aspect of Jamaica's debt reduction is that social indicators gradually improved after 2013. Between 2012 and 2017, the unemployment rate fell from 13.9 percent to 11.6 percent, the household poverty rate declined from 14.4 percent to 13.3 percent, and inequality measured by the Gini index declined from 39.9 percent to 37.5 percent, according to data from the Statistical Institute of Jamaica.[6] These numbers suggest that tight fiscal policy likely had modest negative effects on real activity. Further, improvements in social indicators may have contributed to continued broad-based support for debt reduction and mitigated "fatigue" from running stringent fiscal policy.

BROADER LESSONS What are the lessons from Jamaica's experience? It is tempting to say that Jamaica is a unique case that proves how challenging debt reduction is. The size of the primary surpluses in Jamaica was exceptional, and it is difficult to imagine many other countries sustaining such large surpluses. Yet extreme cases often do carry more general lessons.

6. Statistical Institute of Jamaica, "Living Conditions and Poverty," https://statinja.gov.jm/living_conditions_poverty.aspx.

In my view, Arslanalp, Eichengreen, and Henry point to the two most important lessons: the roles of the fiscal rule and consensus building. Establishing strong fiscal institutions is important for achieving and sustaining debt reduction. Moreover, debt reduction must have buy-in from a broad set of stakeholders. In Jamaica's case, it is difficult to imagine how sustained debt reduction could have been achieved without these factors.

In addition, I would add the following lessons. First, the Jamaica case provides a reminder that fundamental structural and permanent fiscal reforms—rather than temporary measures—are required for sustained debt reductions.

Second, the Jamaican experience suggests that one size does *not* necessarily fit all in terms of how debt reduction is achieved. Debt reduction can be achieved in part by increasing tax revenues, not just by lowering spending. Moreover, in addition to increasing the primary surplus, eliminating "below-the-line" contributors to public debt can be important.

Third, the Jamaica case illustrates that in some cases it takes a crisis, or the threat of a crisis, to implement difficult fiscal adjustments. This is consistent with prior work by Ardagna (2004) and Villafuerte and others (2010). Jamaica's earlier attempts at reducing public debt were not successful. It was only when Jamaica reached the brink of crisis in 2012 that a program was put in place that led to meaningful debt reduction.

Finally, the Jamaica case highlights that it is difficult to predict which fiscal consolidations and debt reductions will work. As noted by several commentators, it was far from obvious ex ante that Jamaica's attempt to reduce its debt would work (Wigglesworth 2020). Jamaica's unlikely success story should thus remind economists and policymakers of the value of humility in making predictions about the effects of large fiscal consolidations.

REFERENCES FOR THE VERNER COMMENT

Alesina, Alberto, and Roberto Perotti. 1995. "Fiscal Expansions and Adjustments in OECD Countries." *Economic Policy* 10, no. 21: 205–48.

Ando, Sakai, Tamon Asonuma, Alexandre Balduino Sollaci, Giovanni Ganelli, Prachi Mishra, Nikhil Patel, and others. 2023. "Coming Down to Earth: How to Tackle Soaring Public Debt." In *World Economic Outlook: A Rocky Recovery.* Washington: International Monetary Fund.

Ardagna, Silvia. 2004. "Fiscal Stabilizations: When Do They Work and Why." *European Economic Review* 48, no. 5: 1047–74.

Asonuma, Tamon, and Christoph Trebesch. 2016. "Sovereign Debt Restructurings: Preemptive or Post-Default." *Journal of the European Economic Association* 14, no. 1: 175–214.

Campos, Camila F. S., Dany Jaimovich, and Ugo Panizza. 2006. "The Unexplained Part of Public Debt." *Emerging Markets Review* 7, no. 3: 228–43.

Cruces, Juan J., and Christoph Trebesch, 2013. "Sovereign Defaults: The Price of Haircuts." *American Economic Journal: Macroeconomics* 5, no. 3: 85–117.

Farhi, Emmanuel, and Iván Werning. 2016. "Fiscal Multipliers: Liquidity Traps and Currency Unions." In *Handbook of Macroeconomics, Volume 2B*, edited by John Taylor and Harold Uhlig. Amsterdam: North-Holland.

Giavazzi, Francesco, and Marco Pagano. 1990. "Can Severe Fiscal Contractions Be Expansionary? Tales of Two Small European Countries." *NBER Macroeconomics Annual* 5: 75–111.

International Monetary Fund (IMF). 2010. *Jamaica: 2009 Article IV Consultation and Request for a Stand-By Arrangement.* Country Report 2010/267. Washington: Author.

International Monetary Fund (IMF). 2014. *Jamaica: Staff Report for the 2014 Article IV Consultation and Fourth Review under the Extended Arrangement under the Extended Fund Facility and Request for Modification of Performance Criteria.* Country Report 2014/169. Washington: Author.

Johnston, Jake. 2015. "Partners in Austerity: Jamaica, the United States and the International Monetary Fund." Washington: Center for Economic and Policy Research (CEPR).

Nickel, Christiane, Philipp Rother, and Lilli Zimmermann. 2010. "Major Public Debt Reductions: Lessons from the Past, Lessons for the Future." Working Paper 1241. Frankfurt am Main: European Central Bank.

Reinhart, Carmen M., Kenneth S. Rogoff, and Miguel A. Savastano. 2003. "Debt Intolerance." *Brookings Papers on Economic Activity*, no. 1: 1–74.

Villafuerte, Mauricio, Cemile Sancak, Jan Gottschalk, S. M. Ali Abbas, Olivier Basdevant, Ricardo Velloso, Fuad Hasanov, Greetje Everaert, Stephanie Eble, and Junhyung Park. 2010. "Strategies for Fiscal Consolidation in the Post-Crisis World." Department Paper 2010/008. Washington: International Monetary Fund.

Wigglesworth, Robin. 2020. "Inside the IMF's Outrageous, Improbably Successful Jamaican Programme (Pt. 1)." *Financial Times*, February 14. https://www.ft.com/content/75274500-6624-40ca-8cfd-67797297c1e9.

GENERAL DISCUSSION Laurence Ball asked for more clarity on where the surpluses came from—whether the burden of taxation is shared differently across the population, and whether the government spending is in a different way relative to other countries. Raghuram Rajan seconded Ball's question and wondered what the government of Jamaica was doing to achieve a shared burden, asking if they raised the taxes on the rich as part of the process. He inquired about more details on how the unemployment rate could come down in the face of tepid economic growth. Peter Henry replied that the surpluses were mainly achieved by spending cuts but included some increases in taxes.

Tristan Reed, thinking back to his coursework in development macroeconomics, noted that institutions were assumed to be the fundamental cause of long-run growth. The authors seemed to show instead that stable institutions give rise to fiscal responsibility. Reed queried the authors about the absence of growth in the case of Jamaica and raised the possibility that this might suggest that other factors, such as geography, deserve more attention as explanatory factors.

Henry discussed different cases of Barbados and Jamaica in relation to the impact of institutions on growth. He explained that the policy choices made by Michael Manley who was elected prime minister of Jamaica in 1972, including raising export barriers and increasing spending as a share of GDP more than twofold, had a devastating impact on the economy, which contracted every year for thirteen years straight starting in 1973. This happened under the same democratic institutional framework as that of Barbados, which did not experience a similar contraction, suggesting policies were to blame for Jamaica's deteriorating economic conditions.

Hoyt Bleakley brought up the benefits that the financial markets have enjoyed, using as an example the fact that Jamaica recently issued debt in its own currency. How large of a wealth effect has this been for the country? Turning to the other side of the capital account, Bleakley noted that government savings had gone up, but he asked whether there had been some sort of credibility effect that had affected private sector external borrowing positively as well.

David Romer queried the authors about the consensus building process that Jamaica went through. He pondered why a populist leader would not be able to achieve what Jamaica achieved, mentioning Chile as an example of a country with a populist leader who has seemingly not yet repudiated their fiscal responsibility.

Kenneth Rogoff wanted to know more about the history of Jamaica, recalling long periods of a poorly run government and asked the authors to provide more details. Steven Kamin similarly was hoping for more details on how the consensus for fiscal consolidation came about. Usually, he argued, it takes an economic crisis to move a society toward accepting budget consolidation. What was the igniting factor in the case of Jamaica? Kamin added that the weak economic growth in the decade or so after might make one wonder whether a consolidation of the magnitude implemented was needed in the first place.

Henry replied that, essentially, Jamaica was in a crisis for about forty years, from around 1972 under Manley to when the debt turnaround happened; and in fact, GDP per capita in Jamaica today is still not where it was in 1972 before Manley came to power. Speaking to the consensus building process, Henry noted that, as explained in more detail in the paper, the first steps toward social partnership were taken with the creation of the Electoral Advisory Committee in 1979 ahead of the 1980 elections. Stakes were incredibly high, Henry explained, and in the run-up to the elections more than eight hundred murders took place. The country was on the brink of complete collapse, and stakeholders realized that something needed to be done—and over the next many years more social partnerships were established. Fast-forward to 2013, the minister of finance and planning, Peter Phillips, embraced the approach of social partnership and, Henry argued, garnered support because people understood the consequences in the absence of cohesion, looking back to 1980. Barry Eichengreen added that the expensive bank bailout, which crippled the government's ability to spend in the late 1990s, was also fresh in the memory of many.

Gian Maria Milesi-Ferretti provided additional context, noting that he was the International Monetary Fund (IMF) reviewer during this time. Milesi-Ferretti explained that because Jamaica had a somewhat strained relationship with the IMF as a result of issues with keeping on track with previous IMF programs, the process of getting the loan to Jamaica was not easy. Milesi-Ferretti also recalled that negotiations with other actors who were needed to secure the loan, including the World Bank and the Inter-American Development Bank, were quite fraught. Ex ante, therefore, the odds of this project succeeding seemed very low. In addition, Milesi-Ferretti said, while we tend to think of $r - g$ as the channel through which growth affects debt, he reminded everyone about the primary balance, pointing out that with weak growth comes lower tax revenues. In sum, this made the effort by policymakers and others involved in finalizing the loan to Jamaica even more impressive. Building on Milesi-Ferretti's accounts, Henry explained that Jamaica had had twelve failed interactions with the IMF before the debt turnaround in 2013, and that in 2012, negotiations with the IMF were virtually nonexistent.

Eichengreen remarked on the discussant Emil Verner's comments about the underlying budget surplus prior to 2013. Eichengreen explained that calculations by the authors suggest that the debt increased between 2006 and 2013, half of which can be attributed to unfavorable $r - g$ and real depreciation, which increased the value of the external dollar denominated debt, and the other half because of hidden government spending on state bodies. The 2013 fiscal rule reform brought that hidden spending to the surface, and it was compressed. This, Eichengreen believed, is compatible with the authors' emphasis on the role of fiscal transparency.

HASSAN AFROUZI
Columbia University

MARINA HALAC
Yale University

KENNETH ROGOFF
Harvard University

PIERRE YARED
Columbia University

Changing Central Bank Pressures and Inflation

ABSTRACT We introduce a simple long-run aggregate demand and supply framework for evaluating long-run inflation. The framework illustrates how exogenous economic and political economy factors generate pressures that, in the presence of central bank discretion, can have an impact on long-run inflation as well as transitions between steady states. We use the analysis to provide a fresh perspective on the forces that drove global inflation downward over the past four decades. We argue that for inflation to remain low and stable in the future, political economy factors, such as strengthened central bank independence or more credible public debt policy, would need to offset the global economic pressures now pushing average long-run inflation upward.

The global increase and persistence in inflation during the past two years has led to much debate regarding the long-term path of inflation. A prevailing view is that inflation levels will not only fall back toward central bank inflation targets, but that they will also on average stay there for the indefinite future. This is certainly true for medium-term official projections: the US Federal Reserve dot plot and the European Central Bank staff project an inflation rate of 2.0 and 1.9 percent in 2026, respectively, essentially at the 2 percent inflation target.[1] Several emerging market central banks

1. See US Federal Open Market Committee (2023) and European Central Bank (2023).

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 2024: 205–241

also project long-term inflation rates very close to their targets (which are generally higher than those of advanced economies).[2]

An alternative albeit less common view, articulated in a recent book by Goodhart and Pradhan (2020), is that persistent structural changes in the global economy will keep future global inflation higher on average than in the past. Put differently, factors that may have previously eased political economy pressures on central banks to inflate—including globalization, demographics, and fiscal restraint (Rogoff 2003)—may be reversing themselves, reawakening a latent inflationary bias. In practice, this upward bias may not necessarily imply a rate of inflation that is continuously above current target levels, but it could materialize in the form of occasional bursts of sharply elevated inflation.

The purpose of this paper is to consider these two views—and the subject of long-run inflation more broadly—using economic theory and data. The economic, social, and geopolitical changes that have taken place over the past few years, especially post-pandemic, are quite striking, and we suggest it is useful to have a framework that encompasses at least some of them. This framework can help us understand the implications of recent developments for the political economy of inflation.

As a starting point, we observe that current approaches to the study of the New Keynesian model assume away political economy issues to such an extent that they are ill-suited for an analysis of long-run inflation. These standard and indeed ubiquitous models typically abstract away from the issue of long-run inflation entirely and consider transitory dynamics around a zero-inflation steady state. There exist some normative models that allow for long-run dynamics, but they too predict that long-run inflation is independent of political economy pressures that interact with the underlying economic environment. More specifically, under the optimal central bank policy with commitment (i.e., assuming the central bank can commit to an infinite sequence of future policies), inflation converges to zero in the long run, a result that holds independently of economic parameters.[3] But what if

2. For example, Banco Central do Brasil (2024) projects an inflation rate of 3.2 percent in 2025, just above its 3 percent inflation target.

3. This result is formalized in Benigno and Woodford (2005) and Schmitt-Grohé and Uribe (2011) in a deterministic economy. Coibion, Gorodnichenko, and Wieland (2012) reach a similar (approximate) conclusion in a numerical analysis of a stochastic economy subject to a zero lower bound.

the issue of commitment is not as thoroughly solved in practice as current consensus posits, and central banks actually use their discretion? What if the past few decades marked an epoch where political economy pressures on central banks to inflate were unusually low? Once that is considered, we show that the standard New Keynesian model gives a perspective on central bank commitment and long-run inflation that goes well beyond the models of Friedman (1968) and Barro and Gordon (1983). In our framework, long-run monetary policy has long-run real economy implications. Moreover, we elucidate how long-run inflation depends on the economic environment both theoretically and quantitatively, and how it evolves dynamically in response to permanent changes in the environment.

Our framework is a heuristic representation of the theoretical model we have analyzed in detail in Afrouzi and others (2023). That model consists of a standard nonlinear New Keynesian economy with sticky-price monopolistically competitive firms, but with a couple of distinct features. First, we depart from the conventional approach of employing a linear approximation in the neighborhood of zero inflation. In doing so, we unmask an important long-run effect of inflation on aggregate demand that gives long-run comparative statics more akin to Tobin (1965) than Friedman (1968). This is not a result of introducing political economy factors; it follows from looking more closely at first-order effects that are obscured in the standard linearization around zero inflation of New Keynesian models. Second, we introduce political economy factors by assuming that the central bank lacks commitment and uses its discretion, with central bank strategies and private sector beliefs that are a function of payoff relevant variables. As Halac and Yared (2022) have shown, this implies that steady-state inflation may be higher than optimal in the New Keynesian model.

Of course, lack of commitment may not matter if the central bank has a strong enough anti-inflation bias. However, we argue in this paper that such a bias cannot be taken for granted looking into the future. In fact, this bias may have been exaggerated as an explanation for the decline in inflation over the past several decades, which was also likely due to a favorable economic environment. In our model, central bank discretion interacts with economic factors, such as globalization, to generate *endogenous* political economy pressures on central banks that drive changes in long-run inflation as well as in the real economy. To account for varying degrees of anti-inflation bias in our framework, we augment the baseline model presented in Afrouzi and others (2023) by considering central bank preferences that might differ from those of households, as in

Rogoff (1985).[4] This extension allows us to also capture the *exogenous* political economy pressures on central banks or changes in institutional design that increase or decrease long-run inflation. The model that we propose is not a normative guide to monetary policy, but a positive model to evaluate long-run inflation given the endogenous and exogenous political economy pressures on central banks.[5]

In our framework, the long-run aggregate supply curve corresponds to the well-known Phillips curve that characterizes New Keynesian models, except that we allow for nonzero steady-state inflation. The long-run aggregate demand curve, which is less familiar, emerges in a nonlinear setting where higher long-run inflation leads to higher price dispersion, and that leads to lower demand. Naturally, if there is perfect long-run indexation to the aggregate price level, this effect would go away. But if one accepts the staggered price setting assumption that plays such a critical role in explaining output and inflation dynamics in the New Keynesian synthesis, then the long-run aggregate demand effect of inflation can be first order.

The long-run aggregate demand and supply curves shift in response to factors that exogenously alter the economic environment and thus, endogenously, change the political economy pressures on central banks (such as structural changes that have an impact on the monopoly power of firms). The curves also shift in response to factors that exogenously change the political economy pressures experienced by central banks directly (namely, factors that affect the stance of monetary policy). The changing central bank pressures lead to changes in long-run inflation and output, which can be quantitatively evaluated.

Our analysis reveals new long-run comparative statics implications of the New Keynesian model and delivers predictions for transitional dynamics across steady states. We show that if deglobalization leads to an increase in firm monopoly power, long-run inflation will increase, and short-run inflation will overshoot its new higher long-run level. That is, if deglobalization were to lead to a new long-run average inflation of 3 percent instead of 2 percent, the short-run inflation rate may temporarily be

4. The central bank that we consider lacks commitment as in the model of Barro and Gordon (1983). However, an implicit type of commitment emerges if society can delegate monetary policy to a central banker whose preferences differ from those of households. The degree to which the central bank values household leisure over consumption reflects its anti-inflation bias.

5. Several papers find evidence that political economy pressures on central banks impact inflation, for example, Weise (2012), Binder (2021), and Drechsel (2024).

much higher. Additionally, since the long-run aggregate demand curve is not vertical in our economy (unlike in the standard model linearized around zero inflation), steady-state output would decline. There are other kinds of shocks that can also affect inflation: for example, a strengthening of central bank independence through an increase in the central bank's anti-inflation bias would lower inflation. In this case, the labor share of income would decline, and monopoly distortions would rise, though with the benefit of a more efficient allocation of resources due to lowered price dispersion. The total impact on real output would depend on whether the decrease in output due to higher monopoly distortions outweighs the increase in output due to reduced price dispersion. Quantitatively, we find in Afrouzi and others (2023) that the second channel dominates and thus output increases in the neighborhood of 2 percent inflation.[6]

The social cost of inflation due to higher price dispersion emerges in our framework whatever the degree of price stickiness or anticipation by firms. The magnitude of the costs of inflation is the subject of some discussion. Nakamura and others (2018) argue that these costs are small; however, work by Christiano (2015), Cavallo, Lippi, and Miyahara (2023), and Afrouzi, Bhattarai, and Wu (2024) suggests that in calibrated economies, even small changes in long-run inflation from 2 percent to 3 or 4 percent can have substantial output costs.[7,8] Accordingly, even small increases in future inflation resulting from global economic pressures could have non-negligible economic costs, and this highlights the importance of counteracting (exogenous) political economy pressures to prevent future inflation from rising. Moreover, if a higher average long-run rate of inflation came about because of infrequent bursts of very high inflation, then the average cost would likely be higher than simply having a steady inflation rate above target. The same principle applies if there is overshooting in the transitions as our model suggests.

6. Note that in the New Keynesian model, inflation does not enter directly into the central bank's objective function as in the ad hoc formulation of the Barro and Gordon (1983) model, but only indirectly through its effect on price dispersion. This effect is significant in our nonlinear New Keynesian economy, even in the long run.

7. The cost of inflation in the input-output production network of Afrouzi, Bhattarai, and Wu (2024) is higher than in one-sector models both because sectors with higher dispersion costs have disproportional effects, and because misallocation in one sector spills over to other sectors.

8. This negative relationship between inflation and output in the long run, which emerges in structural models, is consistent with econometric evidence; see Ascari, Bonomolo, and Haque (2023).

The framework that we present provides a richer narrative explanation for the trend in global inflation over the past four decades, beyond those simply pointing to the advent of increased central bank independence and inflation targeting. Through the lens of the model, the global decline in inflation, which took place beginning in the 1980s and 1990s and which accelerated in the 2000s and early 2010s, can be understood to have been underpinned by rising globalization, the deepening Washington Consensus, and deunionization, which all diminished pressures on central banks to inflate. This view may help explain why inflation declined even in countries where central bank reform was at best limited. For those economies with successful central bank reforms to promote independence and inflation targeting, our predictions are not only consistent with the decline in inflation in response to weakened exogenous political economy pressures, but they are also consistent with the decline in the labor share and rising monopoly profits that were experienced by many of those same economies.[9]

Our framework also provides new perspectives on the path of inflation moving forward. We argue that several global economic trends will, more likely than not, increase pressures on central banks to inflate. These include: deglobalization; rising fiscal pressures due to populism and entitlement spending, the green transition, defense spending, and industrial policy; as well as the concomitant rise in long-term real interest rates. In the face of these global economic trends, central banks no longer as constrained by the zero lower bound (which, in a sense, enhances anti-inflation credibility) may find it increasingly challenging, in political economy terms, to maintain average inflation at current target levels. Temporary periods of elevated inflation—perhaps even as high as post-pandemic—could become more common relative to the past. Thus, in contrast to the three decades ending in 2021, implementing stable and low inflation in future decades may require reforms, such as (even further) strengthened central bank independence or (as unlikely as it may seem) more credible public debt policy, to offset the inflationary pressures on central bankers.

CONNECTION TO DEBATE ON MONETARY NEUTRALITY AND SUPERNEUTRALITY Our work is the first to consider how inflation responds in the long run to persistent economic and political economy pressures on central banks in the New Keynesian model. As such, it connects to much older literature. Since the late 1960s, the dominant paradigm in policy has been Friedman (1968), who posits that money is neutral in the long run. Temporary monetary shocks,

9. See Karabarbounis and Neiman (2014) and De Loecker, Eeckhout, and Unger (2020).

whether to the price level or to inflation, do not have real long-run effects because of an anticipatory channel. Forward-looking firms can only be surprised by monetary shocks temporarily, since they eventually adjust their prices, eroding any of the real effects from temporary monetary shocks.[10]

In the language of the debate on monetary policy of the 1960s and 1970s, money in our model is also neutral in the long run, but it is *not* superneutral; higher steady-state inflation affects real variables. Importantly, this feature does not emerge because firms are myopic; firms in the New Keynesian model set prices in the present in anticipation of future price increases as in Friedman (1968). Rather, because price setting is staggered, long-run inflation affects allocations even in the steady state by changing the long-run dispersion of prices, an effect that is invisible in the standard New Keynesian model linearized around zero inflation. Of course, there are other approaches to modeling the efficiency costs of higher inflation, but the effect that we highlight has long been hiding in plain sight in the most widely used model of central banking. The long-run benchmark of Friedman (1968) coincides with the special case of our framework where firms can index price increases to long-run inflation. In that special case, the stance of monetary policy has no impact on long-run steady-state output, and the long-run aggregate supply curve is vertical.[11]

I. Model of Central Bank Pressures and Long-Run Inflation

We study a simple deterministic environment that is a representation of the model analyzed in Afrouzi and others (2023), but much simplified for expositional purposes. As previously noted, this is a standard New Keynesian model (Clarida, Galí, and Gertler 1999; Woodford 2003; Galí 2015), with monopolistically competitive firms that set prices under Calvo-style rigidity (Calvo 1983). Wages are fully flexible, and households make consumption, labor, and savings decisions. Firms and households optimize their decisions while considering current economic conditions and policies and their expectations of future economic conditions and policies. Critically, however, we

10. The debate on monetary neutrality is far from settled. For example, see recent work by Jordà, Singh, and Taylor (2020) and Ma and Zimmermann (2023).

11. A natural question is whether the effects that we emphasize have quantitative and empirical relevance if the economy is close to the Friedman (1968) benchmark of fully flexible prices. Indeed, for calibrated versions of our model, we find that the economy is close to this benchmark, with an almost vertical long-run aggregate supply curve. However, it is precisely in this case that long-run inflation is most sensitive to small changes in economic and political economy pressures on central banks.

do not follow the literature in assuming that fiscal policy works in the background to provide production subsidies to firms to completely neutralize their monopoly incentives in the steady state.[12] Under such an assumption, typically imposed for tractability, the impact we emphasize of inflation on output and the labor share would become second order. Additionally, as highlighted in our introduction, we do not impose a linearization around zero inflation but instead allow for positive long-run inflation.

We note that in the New Keynesian model, the stance of monetary policy at any given point in time directly maps to a value for the labor share, which is inversely related to the equilibrium level of monopoly distortions (or markups). The more expansionary the monetary policy, the higher the demand, the higher the value of the labor share, and the lower monopoly distortions. Moreover, as we explain below, a constant equilibrium labor share emerges in our model under central bank lack of commitment, with its value being a direct function of exogenous central bank preferences. We use these observations in our analysis to index the choice over monetary policy as a choice over labor share as opposed to inflation or the interest rate. This is for analytical convenience.[13,14] In the Barro and Gordon (1983) framework, of course, monetary policy cannot affect anything real as in Friedman (1968), but that is not the case in the canonical New Keynesian framework, even in the long run.

We next turn to a formal discussion based on a special case of Afrouzi and others (2023), where the analysis collapses to a very simple diagram.

I.A. Steady-State Representation

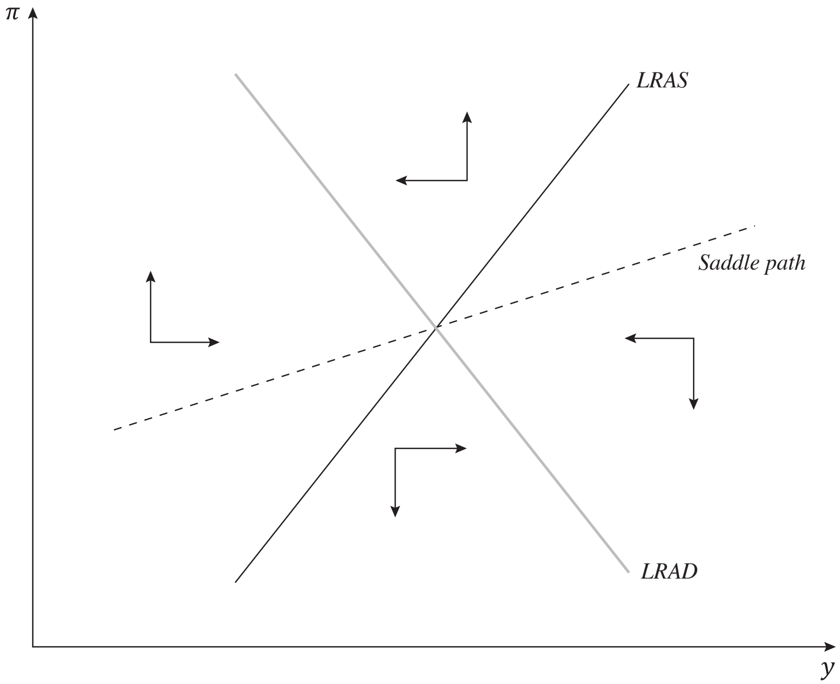
The long-run steady state of the nonlinear model can be represented as corresponding to the intersection of a long-run aggregate supply (LRAS) curve and a long-run aggregate demand (LRAD) curve. As depicted in figure 1, with inflation π on the vertical axis and real (log) output y on the

12. A similar departure is pursued in Benigno and Woodford (2005) and Halac and Yared (2022).

13. The positive relationship between monetary expansion and labor share emerges in a sticky-price and flexible-wage model. A different relationship holds if one instead considers a sticky-wage and flexible-price model; see Galí (2015) for a discussion. Yet, our main results continue to apply in that environment, with comparative statics for union profits that mirror those we show in the text for monopoly profits. These comparative statics are described in our discussion of deunionization in the next section.

14. In the competitive equilibrium of our model, monopoly profits would be zero and the labor share would be one, since the standard New Keynesian model abstracts from capital investment.

Figure 1. Long-Run Aggregate Demand and Supply



Source: Authors' illustration.

horizontal axis, the LRAS curve is upward sloping and the LRAD curve is downward sloping.

Let $\mu > 0$ denote the labor share (as determined by monetary policy) and $\gamma > 0$ the exogenous degree of monopoly power in the economy. For some function f, we can then represent the relationship underlying the upward-sloping LRAS curve by the following equation:

$$\pi = f(\underbrace{y}_{+}, \underbrace{\mu}_{+}, \underbrace{\gamma}_{+}).$$

More specifically, we show in the online appendix that applying an approximation around some low level of long-run inflation $\pi^* > 0$ under certain assumptions, this equation can be written as

$$\pi = \frac{\lambda\left(\rho + \lambda - \pi^*\right)}{\rho - \pi^*}\left(y + \frac{\varphi}{1+\varphi}\log\mu + \log\gamma\right),$$

where the value of inflation π corresponds to the deviation from its long-run value π^*, and analogously for the values of log output y, log labor share $\log \mu$, and log monopoly power $\log \gamma$. Here $\lambda > 0$ denotes the average frequency with which sticky-price firms can change their prices, $\rho > 0$ is the household discount rate, and $\varphi > 1$ is the inverse elasticity of labor supply. Since π^* is low, and in particular lower than the discount rate ρ, this equation yields an LRAS curve that is upward sloping.

The LRAS equation is analogous to the well-known short-run Phillips curve but applied to an economy subject to long-run inflation. The usual logic for the short-run Phillips curve is that firms set higher prices—and therefore there is higher inflation—if firms anticipate higher marginal costs and marginal costs are increasing in real wages, which increase with output y and with the share μ of output paid to workers. Under positive long-run inflation, this relationship captures additional anticipatory effects by price-setting firms. Faced with permanently higher real wages, and therefore larger anticipated absolute changes in future nominal wages (holding fixed the level of inflation), firms face a higher risk of not being able to raise prices in the future as their marginal costs increase. Recall that in the underlying Calvo (1983) model of staggered price setting, individual firms do get to reset their prices, but the timing is uncertain. As such, when a firm gets the chance to change its price in a given period, it will increase it more aggressively in the face of higher real wages, which aggregated across firms results in higher inflation. This anticipatory effect explains why the LRAS curve becomes vertical if prices become fully flexible (or perfectly indexed to long-run inflation) as λ (the average frequency of price adjustment) goes to infinity.

The relationship underlying the downward-sloping LRAD curve is less familiar since it is unique to the nonlinear environment that our approach emphasizes. For some function g, this relationship can be represented by the following equation:

$$\pi = g(\underbrace{y}_{-}, \underbrace{\mu}_{+}).$$

Applying an approximation around some low level of long-run inflation $\pi^* > 0$ as above (see the online appendix for details), this equation can be written as

$$\pi = -\frac{\lambda\left(\lambda - \pi^*\right)}{\pi^*}\left(y - \frac{1}{1+\varphi}\log \mu\right).$$

The LRAD equation emerges because demand y is a negative function of inflation π and a positive function of the labor share μ. Higher inflation leads to higher price dispersion, which contributes to lower demand, with similar goods that are either overpriced or underpriced relative to the average.[15,16] A higher labor share in turn leads to higher demand. This is because the labor share is inversely related to equilibrium monopoly distortions that suppress demand; hence, the higher the labor share, the lower the monopoly distortions and the higher aggregate demand, holding all else fixed. Observe that as π^* approaches zero, the LRAD curve becomes vertical. This illustrates why the standard analysis of the New Keynesian model linearized around zero inflation ignores the effect of inflation on long-run demand.[17]

The long-run steady-state equilibrium corresponds to the intersection of the LRAS and the LRAD curves, as in figure 1. At this intersection, firms optimize prices given the equilibrium level of real wages, and households optimize consumption given the degree of price dispersion.

I.B. Central Bank Preferences

The representation of the steady state described above is general and can flexibly accommodate multiple different frameworks for central bank decision making, including full commitment to zero inflation and lack of commitment. To perform comparative statics and analyze transition dynamics, one must define what the central bank's decision-making framework implies for the equilibrium labor share μ, which is endogenous to monetary policy. As previously noted, in the New Keynesian model, a more expansionary monetary policy stimulates the demand for goods, which stimulates output (which is by assumption demand determined) and employment, thus increasing the labor share.

15. The baseline New Keynesian model abstracts away from efficient sources of price dispersion (for example, stemming from differences in productivity across firms). In the presence of such forces, price dispersion in our framework would be equivalent to dispersion in markups or tax wedges.

16. Demand can be interpreted as the demand from households purchasing from final goods firms or, alternatively, as the demand from final goods firms purchasing from intermediate goods firms.

17. The specific approximation applied above is useful to elucidate the critical nonneutrality that we emphasize outside of zero long-run inflation, but our framework is general and does not require a focus on this special case. It is also worth noting that the LRAD curve in the nonlinear environment becomes upward sloping under deflation, since greater deflation (i.e., more negative inflation) increases price dispersion.

An important case—on which we focus from here on—is one where the labor share μ is kept constant in the short run and long run, independently of economic shocks. We show in Afrouzi and others (2023) that this structure emerges endogenously in a Markov perfect competitive equilibrium in which a central bank that cannot commit makes sequential interest rate decisions. The central bank takes the price-setting process of firms as given and chooses an interest rate that addresses intra-temporal distortions (reflecting both monopoly power and price dispersion). In such an environment, the value of the equilibrium labor share is a direct function of exogenous central bank preferences.

To see this heuristically, let household preferences at any point in time be given by $U(Y) - V(L)$ for increasing functions U and V, where $Y = \exp(y)$ is output, which is equal to consumption, and L is labor, which is inversely related to leisure. The central bank's preferences can be represented by $\mu^*U(Y) - V(L)$, where $\mu^* > 0$ is an exogenous measure of central bank dovishness. In the model analyzed in Afrouzi and others (2023), labor satisfies $L = DY$, where $D \geq 1$ is the degree of price dispersion.[18] The central bank maximizes its static welfare taking the path of prices (and thus dispersion) as given, therefore setting $V'(L)/U'(Y) = \mu^*/D$. Denoting the nominal wage by W and the price level by P, households' intra-temporal optimization sets $V'(L)/U'(Y) = W/P$. Combining these optimality conditions yields

$$\frac{WL}{PY} = \frac{V'(L)L}{U'(Y)Y} = \frac{\mu^* L}{DY} = \mu^*.$$

Thus, the labor share is constant and determined by the exogenous weight the central bank places on consumption over leisure. If the central bank's preferences coincide with those of households, the equilibrium labor share is equal to one because the central bank wishes to undo all equilibrium monopoly distortions; doing so maximizes household welfare conditional on price setting. If instead the central bank has different preferences than households, as in Rogoff (1985), then the equilibrium labor share is some number different from one but still constant over time.[19] A lower labor

18. See equation (12) in Afrouzi and others (2023). Intuitively, higher dispersion implies that more labor is needed to produce a given level of output.

19. We keep the central bank's preferences fixed through time. Halac and Yared (2020) consider equilibrium dynamics when these preferences change over time and are privately known to the central bank, and they show that this can give rise to the presence of persistent high-inflation and low-inflation regimes.

share reflects a higher degree of central bank hawkishness: the central bank places less weight on reducing monopoly distortions and indirectly more weight on reducing inflation.

I.C. Comparative Statics

We illustrate the application of our framework by performing two comparative statics exercises. The first exercise involves an exogenous change in firm monopoly power, which results in *endogenous* political economy pressure on the central bank. The second exercise involves an exogenous change in central bank preferences, which results in *exogenous* political economy pressure on the central bank.

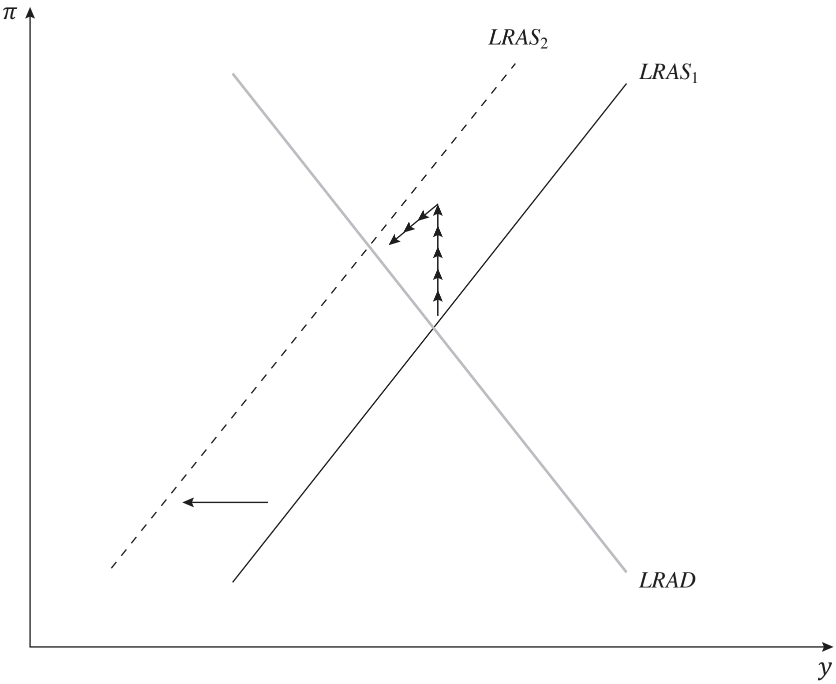
For our first exercise, consider the following change in the economic environment. Suppose that economic factors (e.g., a retreat from globalization) cause the degree of market competition to fall, so that the monopoly power of firms rises permanently. This shock would increase firm monopoly rents in a flexible-price setting. In our sticky-price environment, the shock corresponds to an increase in the parameter γ underlying the LRAS curve and can thus be represented by a leftward shift of this curve in the neighborhood of the steady state, as depicted in figure 2.[20] For every level of real wages, firms with greater monopoly power will now set higher prices, resulting in higher inflation. The economy therefore transitions to a new steady state with higher inflation. Moreover, output is lower since demand responds negatively to higher price dispersion under higher inflation.

Despite the higher monopoly power, monopoly rents stay constant in equilibrium. The reason is that the central bank allows higher inflation in order to prevent the economy from experiencing an equilibrium increase in monopoly rents and decrease in the labor share (which is why the LRAD curve does not shift). The staggered price setting makes it possible for the central bank to lean against exogenous changes in monopoly power and to leave monopoly distortions unchanged. We show in Afrouzi and others (2023) that this is exactly what a central bank without commitment would do. Thus, in the long run, the central bank experiences endogenous political economy pressure due to the changing economic environment, and it is forced to acquiesce to higher inflation.

For our second exercise, consider the following change in the political economy environment. Suppose that the central bank becomes permanently

20. Using a framework where demand elasticities vary with the measure of different varieties, Sbordone (2010) provides a microfoundation for the change in monopoly power that would emerge from an increase in trade.

Figure 2. Effect of Increase in Monopoly Power

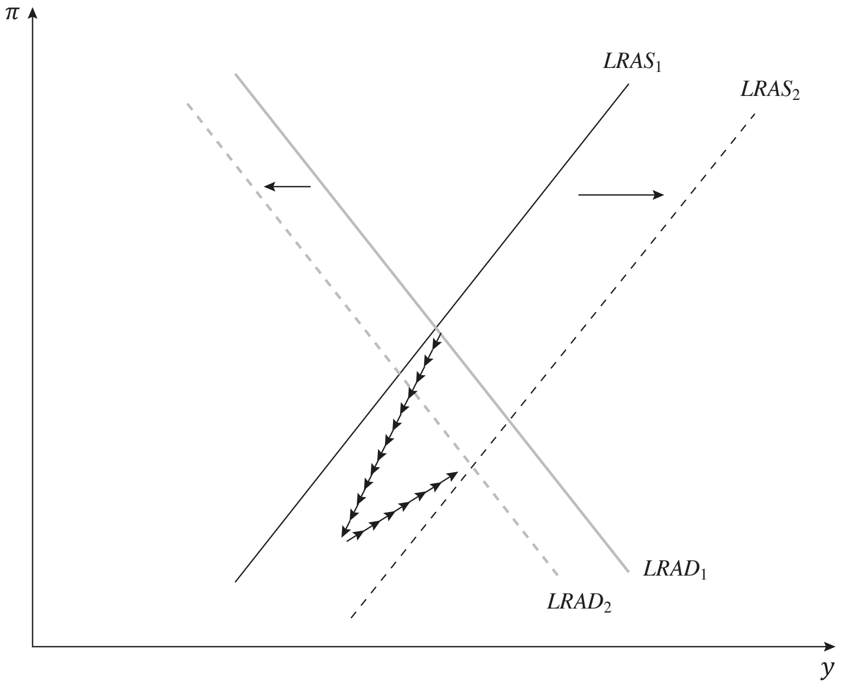


Source: Authors' illustration.

more hawkish, so that it places less weight on reducing monopoly distortions (and indirectly more weight on reducing inflation). This means that the labor share declines, and the degree of equilibrium monopoly distortions rises. As depicted in figure 3, this change can be represented by a rightward shift of the LRAS curve, since there is a lower real wage for every level of output given the lower labor share, and a leftward shift of the LRAD curve, since there are higher monopoly distortions and lower demand for every level of inflation. The economy therefore transitions to a new steady state with lower inflation. The change in output in this case is ambiguous, since it depends on whether the increase in demand due to lower price dispersion exceeds the decrease in demand due to higher monopoly distortions and a lower labor share. In figure 3, the change in output is positive.

This exercise shows that a more hawkish central bank can alter the labor share and the degree of monopoly distortions in the steady state of the economy by changing the level of inflation. At lower levels of inflation, there is

Figure 3. Effect of Increase in Central Bank Hawkishness



Source: Authors' illustration.

less price dispersion and thus less over-hiring by sticky-price firms, yielding a lower labor share and higher monopoly distortions. The example highlights the two key long-run forces in the New Keynesian model. Because it has a single instrument, the most that the central bank can do—for a given degree of price dispersion—is to change long-run demand as a means of changing aggregate (monopoly) distortions, and this is possible because prices are sticky. However, staggered price setting means that the more the central bank alleviates aggregate distortions, the higher the inflation, and the larger the induced variance in the idiosyncratic distortions (price dispersion).

Taken together, the two comparative statics exercises above elucidate what is required in the long run for credible inflation targeting, which is the optimal long-run policy under full commitment. The first comparative static shows that in the face of rising monopoly power, a central bank without commitment will experience pressure to allow higher inflation to stimulate demand to keep monopoly distortions stable. The second comparative static shows that this effect could potentially be offset by a change in exogenous political economy pressure on the central bank, specifically by reducing the degree to which the central bank is concerned with monopoly distortions. In this scenario with two offsetting forces, inflation and price dispersion would remain stable, while output would decline since equilibrium monopoly distortions rise.

I.D. Quantitative Implications

A natural question concerns the quantitative relevance of economic and political economy factors for long-run inflation. We show next that the magnitudes in our framework are significant. Combining the equations underlying the LRAS and LRAD curves presented above, we obtain that the steady-state levels of inflation and output are given by

$$\pi = \frac{(\lambda - \pi^*)(\rho + \lambda - \pi^*)}{\rho} \log(\mu\gamma),$$

$$y = -\frac{\pi^*(\rho + \lambda - \pi^*)}{\rho\lambda} \log(\mu\gamma) + \frac{1}{1+\varphi} \log(\mu).$$

A conventional calibration sets the annual discount rate ρ to around 0.04; the average annual frequency of price changes λ to around 1.2, a 10 percent monthly frequency of price changes (e.g., Nakamura and Steinsson 2008); and the inverse elasticity of labor supply φ to 2.5 (e.g., Chetty and others 2011).

Consider an economy that begins with an inflation rate of 2 percent (i.e., with long-run inflation $\pi^* = 0.02$ and deviation $\pi = 0$). Take a deglobalization scenario in which the country trades less with the rest of the world, causing the degree of market competition to fall and thus the monopoly power of firms γ to rise. The resulting change in inflation depends on the magnitude of the change in γ, which is a function of the extent of deglobalization and the openness of the economy. Edmond, Midrigan, and Xu (2015) estimate the effect of trade openness for Taiwan, and they find that reducing the import share of GDP by 25 percent increases γ by 0.2 percent.[21] A less extreme and more realistic deglobalization scenario would entail a 10 percent decline in import share for an economy that is half as open as Taiwan (which is one of the most open economies in the world). Linear extrapolation for this case translates to a 0.04 percent increase in γ.

21. This measure comes from the change in aggregate markup from increasing the import share of GDP for Taiwan from 30 to 38 in table 3 of Edmond, Midrigan, and Xu (2015).

In our framework, applying a 0.04 percent increase in γ, with parameters taking values as described above, yields an increase in inflation of 1.4 percent and a decrease in output of 0.02 percent. Thus, in this deglobalization scenario, annual inflation rises from 2 percent to 3.4 percent, which is substantial.[22,23] The negative impact on output, in contrast, is small under this calibration, but we anticipate that introducing an input-output structure as in Afrouzi, Bhattarai, and Wu (2024) would increase the magnitude of this effect. In fact, their work shows that such an extension of the New Keynesian model increases the output cost of inflation by tenfold to twentyfold.

I.E. Transition Dynamics

Our framework can also be used to study transition dynamics. To the right of the LRAS curve in figure 1, output and therefore real wages exceed the steady-state level. This means that inflation—which captures expectations of the path of future real wages—falls in this region. The opposite is true to the left of the LRAS curve. Analogously, to the right of the LRAD curve, inflation exceeds the steady-state level. This means that dispersion—which captures the historical path of inflation—rises in this region, and therefore demand falls, since households demand fewer goods when there is rising variance in prices. The opposite is true to the left of the LRAD curve, where demand increases. These flows are depicted by the arrows in the different regions in figure 1 and putting them together allows us to define a saddle path around the steady state. As also shown in figure 1, the saddle path yields transition dynamics that admit positive co-movement between inflation and output.

Consider how an economy transitions from an initial steady state to a new higher-inflation steady state in response to a change in the economic environment. For concreteness, take our first comparative static exercise, depicted in figure 2, where firm monopoly power exogenously increases. Starting from the initial steady state, inflation must immediately jump to the new saddle path in response to the shock, and it must then fall along the saddle path toward its new steady-state level. That is, the transition must feature inflation overshooting.

The logic for overshooting is as follows. The initial jump in inflation is a direct response to the exogenous increase in monopoly power. This rise

22. In the online appendix, we show that the quantitative effects from the approximate linearized model are in line with those of the nonlinear model.

23. We note that if the frequency of price adjustment λ responds positively to equilibrium inflation (as would be the case in a menu cost model, for example), then the quantitative impact on inflation would be even larger.

in inflation, however, only leads to a gradual increase in price dispersion. Along the path toward a new higher-dispersion steady state, demand declines, and this is reflected in a downward path for output and real wages. In turn, this implies a downward path for the marginal costs faced by firms, which explains the declining level of inflation toward the new steady state. Consequently, if deglobalization leads to greater monopoly power and higher long-run inflation, the short-run spike in inflation can be much greater.

For another example, take our second comparative static exercise, where the central bank becomes permanently more hawkish, and where the long-run steady state of the economy shifts to a higher level of output, as depicted in figure 3. Starting from the initial steady state, output immediately jumps down and then rises toward its new higher steady-state level. The immediate downward jump reflects higher monopoly distortions, whereas the eventual output increase reflects lower price dispersion. The path of inflation follows by analogous reasoning to the previous example, since real wages and output rise along the equilibrium path after the initial downward jump.

The opposite transition dynamics would hold if the central bank instead became permanently more dovish. In that case, output and inflation would immediately jump upward and then decline toward a lower steady-state output and higher steady-state inflation level.

II. Historical Inflation through the Lens of the Model

We apply our framework to provide a fresh perspective on the economic and political economy forces that drove global inflation over the past four decades.

II.A. Empirical Evidence

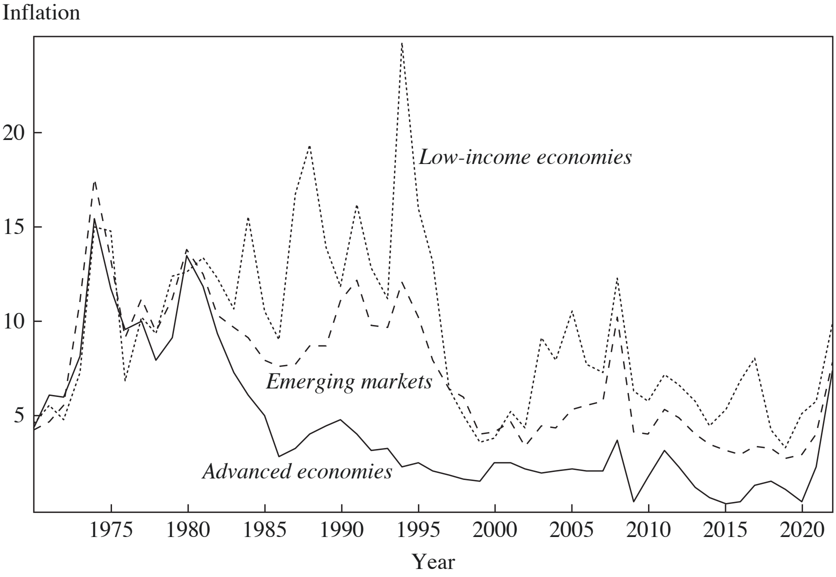
Figure 4 depicts inflation across the world for three different country groups over the period 1970–2022.[24]

The figure illustrates the global decline in inflation that peaked in advanced economies in the early 1980s, and globally in the early 1990s. The rapid decline in advanced economy inflation during the 1980s, after the high-inflation experience of the 1970s, has been widely studied in the literature (e.g., Sargent 2001; Primiceri 2006; Bianchi 2013; Nelson 2022).

Less studied, but equally salient is the global decline in inflation in the 1990s, 2000s, and 2010s. This decline had an impact on all country groups

24. The inflation rates are for a balanced panel and correspond to the median. Similar patterns are observed if we instead use an unbalanced panel or GDP-weighted measures.

Figure 4. Inflation across the World



Source: Data from Ha, Kose, and Ohnsorge (2023) at https://www.worldbank.org/en/research/brief/inflation-database.

across each decade, including low-income countries, albeit their decline was not as steady and smooth as in advanced economies. Inflation in the emerging market and low-income country groups fell to around 5 percent around the turn of the century but then rose to over 10 percent even before the global financial crisis, falling again in the mid-2010s, and rising sharply again after the pandemic. The ebbs and flows in the 21st century do not necessarily reflect formal changes in central bank independence. Indeed, many low-income countries experienced only limited central bank reform. As an example, annual inflation in Uganda decreased from an average of 17.8 percent in the 1990s to an average of 6.4 and 6.6 percent in the 2000s and 2010s, respectively. During that time, various measures of central bank independence for Uganda stayed the same or even deteriorated.[25] For emerging markets, average inflation declined from 5.3 percent in the 2000s to 3.8 percent in the 2010s, and this occurred even though measures

25. This is based on the data in Romelli (2022), which measures various aspects of central bank independence, extending the work by Cukierman, Webb, and Neyapti (1992).

of central bank governance in these countries deteriorated after the global financial crisis of 2008 (e.g., Bordo and Siklos 2021).

Through the lens of our model, the long-run decline in global inflation can be viewed in part as the result of three global forces that changed endogenous political economy pressures on central banks: rising globalization, the deepening Washington Consensus, and deunionization. The decline in inflation also clearly reflects exogenous political economy pressures, as reflected by many successful central bank reforms that promoted independence and inflation targeting. We address each of these phenomena and their implications in our model separately.

II.B. Globalization

Between 1970 and 2007, global trade as a proportion of global GDP increased from 25 percent to 59 percent.[26] As is well known, the era of hyperglobalization was a consequence of containerization, which dramatically diminished the cost of shipping. It was also driven by the reduction in tariff barriers and proliferation of trade agreements and dispute resolution processes, marked by major landmarks, such as the creation of the European Union in 1993 and the accession of China into the World Trade Organization in 2001. Financial globalization, unleashed by the relaxation of capital controls, further facilitated trade globalization by allowing for trade imbalances to form, while also fostering the flow of foreign direct investment. Between 1970 and 2007, foreign direct investment as a share of global GDP increased from 0.5 percent to 5.3 percent.[27]

In our framework, the increase in global competition translates to a reduction in firm monopoly power γ. In fact, this is supported by empirical evidence; for example, Bloom, Draca, and Van Reenen (2016) find lower prices and profitability for European firms more exposed to China's entry into the World Trade Organization relative to those that were less exposed.[28] The reduction in γ in our model has the effect of shifting the

26. World Bank, "Trade (% of GDP)," https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS.

27. World Bank, "Foreign Direct Investment, Net Inflows (% of GDP)," https://data.worldbank.org/indicator/BX.KLT.DINV.WD.GD.ZS.

28. Edmond, Midrigan, and Xu (2015) find similar results in a model calibrated to Taiwanese data. De Loecker and others (2016) find that trade liberalization in India led to a decrease in prices but also a further decrease in *equilibrium* marginal costs, therefore resulting in higher equilibrium markups. However, to the extent that *efficient* marginal costs are independent of trade, their findings imply a decrease in markups relative to the latter, and thus a decrease in the monopoly distortions that drive the central bank's incentive to inflate in our model.

LRAS curve to the right, which means that the economy transitions to a new steady state with lower inflation and higher output (due to lower price dispersion), holding fixed the level of central bank hawkishness. Globalization thus reduces the endogenous political economy pressure on the central bank to inflate, resulting in lower inflation.

II.C. Washington Consensus

A second force to consider is the proliferation of the Washington Consensus, a term that refers to a set of widely adopted programs of market liberalization, privatization, and fiscal discipline. This program of reforms was implemented across countries in the 1980s, 1990s, and 2000s, often with the support of international institutions.

Between 1985 and 2001, the fraction of countries classified as market-oriented increased from 30 percent to almost 80 percent (Buera, Monge-Naranjo, and Primiceri 2011). In Latin America, for example, more than eight hundred public enterprises were privatized between 1988 and 1997 (Aninat 2000). The effects of market liberalization and privatization in our framework are isomorphic to the effects of globalization that we discussed above. These reforms reduce firm monopoly power γ, which results in lower inflation and higher output, holding fixed the level of central bank hawkishness.

On the fiscal side, the process of reform led to a decline of the public debt-to-GDP ratio in emerging markets from a peak of 68 percent in 2002 to 46 percent by 2015. In low-income countries, the ratio declined from a peak of 99 percent in 1994 to 48 percent by 2015.[29] Of course, these patterns are in direct contrast to the experience of advanced economies, which witnessed a secular long-term increase in public debt from the mid-1970s onward (Yared 2019).

To evaluate the effects of reduced fiscal pressures in emerging markets and low-income countries, we can consider an extension of our framework that incorporates fiscal objectives for the central bank, as in Schreger, Yared, and Zaratiegui (2023). Their work shows that the central bank responds to diminished fiscal pressures with lower desired monetary stimulus. In particular, the lower the inherited public debt, the lower the pressure on the central bank to use inflation to devalue that debt to mitigate the economic cost of debt repayment. Moreover, the lower the deficit, the lower the pressure on the central bank to stimulate the economy to reduce the real interest

29. International Monetary Fund, "Debt (% of GDP)," https://www.imf.org/external/datamapper/DEBT1@DEBT/FAD_G20Adv/FAD_G20Emg/FAD_LIC.

rate and the cost of issuing new debt.[30] Through these two channels, lower fiscal pressures reduce the endogenous pressures on the central bank to inflate. In our model, this translates to a lower labor share μ, which has the effect of shifting the LRAS curve to the right and the LRAD curve to the left. The result is a transition to a new steady state with lower inflation and lower price dispersion.

This discussion suggests that, in principle, there is a strong political economy mechanism for fiscal policy to influence inflation through its effect on central bank incentives. While this mechanism differs from the fiscal theory of the price level—which argues for a direct effect of fiscal policy on inflation independently of monetary policy—it is consistent with the empirical correlation between deficits and inflation that supports that theory (e.g., Barro and Bianchi 2023; Cochrane 2023).[31]

II.D. Deunionization

A third important force is deunionization, particularly in advanced economies. In the United States, the fraction of households in trade unions declined from 22 percent in 1980 to 11 percent by 2010. Out of twenty-four advanced economies with available data, twenty experienced a reduction in unionization rates over this period, including countries like Germany where unionization rates have been historically high.[32]

To evaluate the effects of deunionization in our framework, we can consider an analogous model to ours but allowing for labor market power instead of firm market power. Specifically, we can take a model with sticky wages and flexible prices (instead of sticky prices and flexible wages), again accounting for nonlinearities.[33] The LRAS and LRAD curves are defined analogously to our previous analysis, with *wage* inflation (which equals price inflation) on the vertical axis and real (log) output on the horizontal axis. The LRAS curve corresponds to a steady-state wage Phillips curve. The LRAD curve corresponds to a firm labor demand curve—demand

30. While the steady-state real interest rate is exogenous in Afrouzi and others (2023), the New Keynesian overlapping generations framework of Aguiar, Amador, and Arellano (2023) has a steady-state real interest rate that is endogenous to monetary policy, with higher money growth reducing this rate and expanding fiscal capacity.

31. This mechanism is consistent with the argument in Chari, Henry, and Reyes (2021) that chronic budget deficits in Latin America were a root cause of the region's high inflation levels in the 1980s and early 1990s.

32. OECD, "How Do Collective Bargaining Systems and Workers' Voice Arrangements Compare across OECD and EU Countries?" https://www.oecd.org/employment/ictwss-database.htm.

33. See Galí (2015) for an exposition.

declines as wage inflation and wage dispersion rise. Firms make zero profits in this model, while unions make positive profits. Rather than being indexed by the level of monopoly profits, the stance of monetary policy is now indexed by the level of union profits, with a more hawkish monetary policy corresponding to higher union profits (and therefore larger equilibrium intra-temporal distortions, which imply lower equilibrium inflation, as in the sticky-price, flexible-wage model).

In this extended framework, a decrease in labor market power can be depicted as a rightward shift of the LRAS curve, since unions then set lower wages for every level of output. The result is lower wage inflation and therefore lower price inflation, together with higher output due to lower wage dispersion. Lower labor market power thus reduces the central bank's endogenous political economy pressures to inflate, resulting in lower inflation.[34]

II.E. Central Bank Reform

Central bank reforms across the world—often made in concert with international institutions—are also an important cause of the decline in global inflation. Starting in the mid-1980s, when academic research began to emphasize the potential effectiveness of central bank independence in controlling high inflation, one country after another instituted reforms. Substantially greater independence allowed central banks to adopt inflation targeting mandates, which served as a tool to further enhance their independence (e.g., Bernanke and Mishkin 1997; Bernanke and others 1999). Increased transparency has also played a central role.

Based on data on legislative reforms of central bank charters, 80 out of the 113 central banks with available data experienced an improvement in independence between 1990 and 2010 (Romelli 2022). Dincer, Eichengreen, and Geraats (2022) analyze measures of central bank transparency, and they find that in 100 out of 112 countries with available data transparency increased between 1998 and 2019. Along with these reforms, sixty central banks adopted inflation targeting.[35] Early adopters were central banks in advanced economies, like those in New Zealand, Canada, and the United Kingdom, while more recent adopters included emerging economies such as India and Russia.

34. Stansbury and Summers (2020) argue that declining worker power created disinflationary pressure in the United States over 1982–2016.

35. This includes the nineteen countries in the eurozone plus forty-one other countries classified by the International Monetary Fund (2020) as inflation targeters.

An indirect factor which interacted with these central bank reforms, particularly in advanced economies, is the emergence of the zero lower bound on interest rates, first in Japan in the late 1990s and then in advanced economies in the aftermath of the global financial crisis of 2008 (although the issue had already come into view after the bursting of the tech bubble in 2001). In equilibrium, the expectation that the central bank's hands are sometimes tied serves to lower long-run average inflation expectations.[36] From this perspective, the effect of the zero lower bound is the same as having a central bank with a more hawkish tilt, though this is only on average, since outside zero lower bound episodes, inflation will be higher than under a hawkish central bank.

Through the lens of our model, central bank reforms along with the constraints of the zero lower bound can be studied as an exogenous increase in central bank hawkishness. As previously described, this translates to a lower labor share μ in our framework, shifting the LRAS curve to the right and the LRAD curve to the left, and therefore resulting in lower inflation and lower price dispersion. Observe further that a consequence of these central bank reforms is higher monopoly distortions along with the lower labor share.

II.F. Taking Stock

We have argued that the global decline in inflation over multiple decades can be viewed as resulting from the confluence of exogenous economic and political economy forces that jointly reduced central bank pressures to pursue expansionary monetary policy. We believe that, while very important, central bank reforms on their own cannot explain many of the empirical patterns in figure 4. For example, they cannot explain why inflation declined in countries that experienced little improvement (or even a deterioration) in central bank governance, or why inflation declined in economies that were far away from the zero lower bound. It appears that global economic trends also played a key role by reducing the endogenous political economy pressures on central banks to inflate.

Our view is further supported by the significant heterogeneity in the inflation experience across countries, which cannot be explained by exogenous political economy pressures alone. As depicted in figure 4, low-income countries have on average much higher inflation rates than emerging

36. As an illustration, the Markov perfect equilibrium in a linearized economy, as in Halac and Yared (2022), with the addition of a zero lower bound would predict lower average inflation as a result of a more binding zero lower bound.

markets, which in turn have higher inflation rates than advanced economies. Of course, this reflects in part differences in central bank governance across these country groups, and it is consistent with econometric evidence that finds a negative correlation between long-run inflation and central bank independence across countries.[37] What is clear, however, is that there continues to be heterogeneity in long-run inflation rates even after controlling for central bank independence, and this remaining heterogeneity can be explained by other economic factors. For example, Campillo and Miron (1997) find that countries that are more open to trade or have lower public debts have lower inflation rates. These cross-country findings are consistent with our framework in which economic factors affect the endogenous political economy pressures experienced by central banks.[38]

III. Future Inflation through the Lens of the Model

Figure 4 shows that in the mid to late 2010s, inflation in every country group reached a forty-year trough prior to the post-pandemic inflation spike. For advanced economies, that trough occurred in 2015, with an inflation rate of 0.40 percent; for emerging markets, it occurred in 2019, with an inflation rate of 2.79 percent; for low-income countries, it occurred also in 2019, with an inflation rate of 3.35 percent.

An important, natural question is whether global inflation in the 2020s will return to the levels of the 2010s or instead increase to the levels of the 2000s or even the 1990s. Our model tells us that the answer depends on the likely evolution of economic and political economy forces. We believe that several persistent global economic trends that accelerated during the pandemic—some of which are reversing the decades-old developments described in the previous section—will likely increase central bank pressures to inflate. This means that implementing stable and low inflation in the future may require even further strengthened central bank independence to counteract these endogenous political economy pressures. We describe the sources of the new pressures in this section.

37. See Berger, De Haan, and Eijffinger (2001) for a survey.

38. Note further that using our framework, we can study the relationship between the labor share and inflation. It is well known that the labor share has declined in many countries over decades (e.g., Karabarbounis and Neiman 2014), and there is some debate as to whether this trend reflects a rise in monopoly power (Karabarbounis and Neiman 2018; Philippon 2019; De Loecker, Eeckhout, and Unger 2020) or other factors like a decline in union power (Elsby, Hobijn, and Şahin 2013). Through the lens of our model, the decrease in the labor share and in inflation can be viewed as a joint consequence of a reduction in labor market power or an increase in central bank hawkishness.

III.A. Reversal of Globalization

The globalization trends of prior decades have been reversing since the end of the global financial crisis of 2008. Trade as a proportion of global GDP stopped increasing after hitting a peak at 61 percent in 2008, and it has since declined to 57 percent in 2021.[39] Foreign direct investment as a share of global GDP peaked at 5.3 percent in 2007 and has since declined to 2.2 percent in 2021.[40] In addition to these absolute changes, international flows have also become more fragmented. For example, trade and capital flows in the aftermath of Russia's invasion of Ukraine in 2022 have segmented along geopolitical lines, a development especially costly for Europe, which depends on geopolitically nonaligned countries for trade (Gopinath 2023). Firm-level network data further indicate that global value chains, particularly those that connect to China, have lengthened over the last two years (Qiu, Shin, and Zhang 2023), suggesting an increase in trade costs.

These developments have two main causes, which are likely to remain dominant in the future. The first cause is the application of protectionist trade policies across the world after the global financial crisis, a process that accelerated after the 2020 pandemic. This resulted in a transition from hyperglobalization prior to the global financial crisis to "slowbalization" (Aiyar and Ilyina 2023; Goldberg and Reed 2023), which occurred in large part because of a political backlash against free trade. Restrictions on international flows have been widely applied across countries and go beyond the more salient case of Brexit in 2016 or the US-China trade war beginning in 2018. The number of trade restrictions imposed annually worldwide increased from under 500 in 2010 to around 1,000 in 2018 to almost 3,000 in 2022 (International Monetary Fund 2023a). In addition, the number of countries introducing or expanding security-related screening mechanisms for foreign direct investment increased from under 10 for every year between 1995 and 2019 to 22 in 2020, 17 in 2021, and 14 in 2022 (Guazzini, Leskova, and Meloni 2023).

The second cause for these global developments is the rise in geopolitical tensions. These increased following the Russian invasion of Ukraine in 2022. In response, the United States, European Union, and their allies applied trade and financial sanctions on Russia, resulting in a rerouting of

39. World Bank, "Trade (% of GDP)," https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS.

40. World Bank, "Foreign Direct Investment, Net Inflows (% of GDP)," https://data.worldbank.org/indicator/BX.KLT.DINV.WD.GD.ZS.

global flows. In addition, the Israel-Gaza war in 2023 following the Hamas attack on Israel and the expansion of the conflict to the broader region has led to attacks on commercial vessels in the Red Sea, leading to further disruptions in global trade.

This is a fast-evolving situation given the rising number of measures distorting trade and investment, and the rising geopolitical risk (Caldara and Iacoviello 2022).[41] If countries pursue protectionist policies and do not de-escalate geopolitical tensions in the coming years, then the slowdown in globalization, the rising fragmentation of global flows, and the lengthening of supply chains will also persist. The result is lower global competition and higher firm monopoly power. In our framework, this is reflected in an increase in the monopoly power parameter γ, which shifts the LRAS curve to the left (holding the level of central bank hawkishness fixed) and results in higher inflation and lower output (due to higher price dispersion).[42] Thus, through this channel, a reversal in globalization trends increases the endogenous political economy pressures on the central bank to inflate.[43]

III.B. Rising Fiscal Pressures

A second important trend is increasing global fiscal pressures. The International Monetary Fund projects higher government debt to GDP in the 2020s relative to the 2010s for all country groups: advanced economies, emerging markets, and low-income countries (International Monetary Fund 2023b). Debt overhang from pandemic-era government spending combined with high interest rates is a common driver of this trend, but it is not the only one; rising government primary deficits are also to blame. In advanced economies, the primary deficit as a percentage of GDP is projected to increase from a pre-pandemic (2014–2019) average of 1.2 percent to a post-pandemic (2023–2028) average of 2.2 percent. For emerging markets, the increase is from 2.1 percent to 2.9 percent.

The fiscal pressures for advanced economies largely reflect the acceleration of the aging of the population and the resulting expansion of entitlement spending without commensurable revenue increases (Yared 2019). In the United States, for example, the Congressional Budget Office forecasts

41. See the Global Trade Alert Database, https://www.globaltradealert.org/.

42. Note that lower global competition could also result in an increase in labor market power in a sticky-wage and flexible-price model such as the one we described in the previous section, leading to the same comparative static for inflation.

43. There is direct evidence for a long-run correlation between global geopolitical risk and global inflation that is consistent with this channel (e.g., Caldara and others 2023).

that between 2023 and 2033, Social Security spending will increase from 5.1 percent to 6.0 percent of GDP. Outlays for major health programs will increase from 5.8 percent to 6.6 percent of GDP over that time, with around 25 percent of the increase due to aging (CBO 2023). For emerging markets, the fiscal pressures reflect increasing government spending, particularly in the two largest emerging market economies, China and India.

There are reasons to think that current fiscal forecasts—which only incorporate current policies but not likely changes to future policies—may be too optimistic. For example, more than 140 countries, including the United States, countries in the European Union, China, and India, have set net zero carbon emissions targets. According to simulations by the International Monetary Fund, the government spending policies required to achieve net zero emissions midcentury would increase the forecasted government debt-to-GDP ratio by 10 to 15 percent in advanced economies and 15 percent in emerging markets (International Monetary Fund 2023b).

Similarly, economic forecasts do not adequately account for a potential continuation or escalation of geopolitical tensions, which would likely result in additional defense spending. The Congressional Budget Office forecast—which already predicts a stark trajectory for US debt—assumes that US defense spending as a share of GDP will decline from 3.2 percent in 2023 to 2.7 percent in 2033.[44] Should geopolitical tensions persist, a more realistic forecast would account for the possibility that US defense spending returns to levels closer to those reached during the Cold War, which averaged nearly 7 percent of GDP between 1960 and 1991.[45]

A further consideration for fiscal forecasts is the continuing expansion of industrial policy. These policies—which seek to reorient an economy's resources and production toward national strategic goals—are not just confined to the 2022 CHIPS and Science Act or the 2022 Inflation Reduction Act in the United States; they represent a longer-term global trend. Juhász and others (2023) analyze the text of commercial policies across the world, and they find that the share of policies that can be classified as industrial policies increased from 20 percent in the early 2010s to nearly 50 percent by 2019. Juhász, Lane, and Rodrik (2023) find that the fiscal impact of these policies can range from 0.3 to 0.7 percent of GDP annually.

As described in the previous section, increased fiscal pressures result in higher monetary stimulus: the central bank experiences pressure to use

44. This number is imputed under the Congressional Budget Office's assumption that the proportion of discretionary spending accounted for by defense remains stable at 49 percent.

45. Stockholm International Peace Research Institute, "SIPRI Military Expenditure Database," https://www.sipri.org/databases/milex.

inflation to devalue outstanding public debt and to stimulate the economy to reduce the real interest rate and the cost of issuing new debt. Through both channels, higher fiscal pressures increase the endogenous pressures on the central bank to inflate. This translates to a higher labor share μ in our model, shifting the LRAS curve to the left and the LRAD curve to the right, and thus resulting in higher inflation and higher price dispersion.[46]

III.C. Unshackling from the Zero Lower Bound

A third development having an impact on central banks is the likely upward trajectory in long-term real interest rates back to their centuries-old trend, after deviating from that trend substantially in the aftermath of the global financial crisis (Rogoff, Rossi, and Schmelzing 2022). This change would result in higher nominal interest rates (holding expected inflation constant), thus moving the economy further away from the zero lower bound. This would diminish the de facto hawkish tilt that the zero lower bound imposes on central banks, since then interest rate increases can be more easily offset by interest rate decreases on average. In our model, this translates to a higher labor share μ, shifting the LRAS curve to the left and the LRAD curve to the right, and thus resulting in higher inflation and higher price dispersion.[47]

III.D. Assessment

We have described several forces that would increase the endogenous political economy pressures on central banks to inflate in the 2020s relative to previous decades.[48] Of course, there are many reasons for why our assessment could be wrong.

First, we must accept the possibility that the economic forces we have highlighted may not persist. Perhaps global geopolitical tensions de-escalate,

46. Moreover, to the extent that these fiscal pressures come hand in hand with economic distortions that raise the market power of firms, they can increase inflation by shifting the LRAS curve even further to the left. Consistent with our analysis, Del Negro, di Giovanni, and Dogra (2023) find that green policies change the trade-offs for central banks and can increase their incentives to stimulate the economy.

47. Because they translate to higher interest costs for the government, higher long-run real interest rates also translate to higher fiscal pressures, which further increase central bank pressures to inflate.

48. We note that this list is not exhaustive, and others like Goodhart and Pradhan (2020) would point to demographic pressures as an additional force driving long-run inflation upward. Through the lens of our framework, we can articulate their conjecture as an argument that aging should raise labor scarcity and increase labor market power, thus reversing the impact of deunionization described in the previous section.

with a resumption of long-term globalization trends and a reinvigoration of the Washington Consensus. Perhaps, and more realistically, these forces do persist, but there are other forces that keep inflation from rising. One possibility is that the zero lower bound continues to constrain central banks because of other pressures—for example, demographic ones—keeping long-term real interest rates suppressed. Under this scenario, central banks would find themselves powerless to raise inflation despite endogenous political economy pressures on them to pursue expansionary monetary policy. Another possibility, as some currently argue, is that artificial intelligence and other new technologies will act as a disinflationary force (Klebnikov 2023). In our framework, such technologies would need to reduce monopoly power or alleviate fiscal pressures by boosting economic growth; of course, this force would have to be strong enough to counteract other inflationary forces we have highlighted.

Second, we must also accept the possibility that even if the economic forces driving inflation upward persist, they could be counteracted by exogenous political economy pressures. These would take the form of a renewed push for promoting central bank independence across the world, with a strengthened commitment to containing inflation as opposed to other goals. These efforts could be potentially supported by the public backlash against the inflation surge of 2022 (Stantcheva 2024). Now, a critical difference relative to the past thirty years of central bank reforms is that these efforts would need to work in opposition to, not in tandem with, the endogenous political economy pressures on central banks. Moreover, we should keep in mind that elected politicians have historically always interfered with central bank operations, and the concept of central bank independence is a relatively new one. This reality suggests that success would be more likely if central bank reforms were buttressed by efforts at putting public debts on a sustainable path, potentially through the application of stricter fiscal rules (Yared 2019; Dynan 2023). There are signs of hope: despite the rise of populist policies around the world and the rhetorical attacks on the Washington Consensus, many emerging markets have maintained the key elements of past reforms, placing a premium on macroeconomic stability; this has contributed to their surprising resilience and contained inflation in the face of the major shocks of the past decade and a half (Rogoff 2023). Through the lens of our model, monetary and fiscal reform translate to a lower labor share μ, shifting the LRAS curve to the right and the LRAD curve to the left, and thus resulting in lower inflation and lower price dispersion.

IV. Conclusion

We have presented a simple long-run aggregate demand and supply framework for studying long-run inflation and transition dynamics. Using this framework, we provided a fresh perspective on the economic and political economy forces that drove global inflation downward over the past four decades. Our analysis highlights the underlying reasons why maintaining low and stable inflation may be challenging in the coming decade, and why a strengthening of central bank independence combined with a more credible public debt policy is likely needed to offset the global economic pressures pushing long-run inflation upward. It is worth noting that if political economy pressures do result in higher average inflation, this will likely come in the form of occasional bursts of inflation, such as after the pandemic, rather than an inflation rate that continuously exceeds the target.

Because it is based on the familiar and most widely used model of central banking, we believe that our framework is a useful first step for evaluating the causes and consequences of changes in long-run inflation. The framework clarifies that long-run inflation interacts in important ways with market power to influence aggregate (monopoly) distortions as well as idiosyncratic distortions (price dispersion) in the economy. Assessing what this observation implies more generally—that is, beyond the benchmark single-agent, one-sector, closed-economy New Keynesian model—both for central bank incentives and for the long-run real effects of monetary policy, is an important next step.

ACKNOWLEDGMENTS For helpful comments, we are grateful to Janice Eberly, Brett House, Donald Kohn, Jennifer La'O, Christian Moser, Jesse Schreger, Jón Steinsson, Silvana Tenreyro, Shang-Jin Wei, and participants at the *Brookings Papers on Economic Activity* Spring 2024 Conference. We thank Krishna Kamepalli for excellent research assistance.

References

Afrouzi, Hassan, Saroj Bhattarai, and Edson Wu. 2024. "The Welfare Cost of Inflation in Production Networks." Working Paper. https://afrouzi.com/#working-papers.

Afrouzi, Hassan, Marina Halac, Kenneth S. Rogoff, and Pierre Yared. 2023. "Monetary Policy without Commitment." Working Paper 31207. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w31207.

Aguiar, Mark A., Manuel Amador, and Cristina Arellano. 2023. "Pareto Improving Fiscal and Monetary Policies: Samuelson in the New Keynesian Model." Working Paper 31297. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w31297.

Aiyar, Shekhar, and Anna Ilyina. 2023. "Charting Globalization's Turn to Slowbalization after the Global Financial Crisis." IMF Blog, February 8. https://www.imf.org/en/Blogs/Articles/2023/02/08/charting-globalizations-turn-to-slowbalization-after-global-financial-crisis.

Aninat, Eduardo. 2000. "Growth and Stability in Latin America and the Caribbean: Challenges for the Epoch of Globalization." 14th Eric Williams Memorial Lecture, International Monetary Fund, May 26. https://www.imf.org/en/News/Articles/2015/09/28/04/53/sp052600.

Ascari, Guido, Paolo Bonomolo, and Qazi Haque. 2023. "The Long-Run Phillips Curve Is . . . A Curve." Working Paper 789. Amsterdam: De Nederlandsche Bank. https://www.dnb.nl/media/vfcmbaat/working_paper_no-789.pdf.

Banco Central do Brasil. 2024. *Inflation Report, December 2023*. Vol. 25, edition 4. Brasília: Central Bank of Brazil. https://www.bcb.gov.br/en/publications/inflationreport.

Barro, Robert J., and Francesco Bianchi. 2023. "Fiscal Influences on Inflation in OECD Countries, 2020–2022." Working Paper 31838. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w31838.

Barro, Robert J., and David B. Gordon. 1983. "Rules, Discretion and Reputation in a Model of Monetary Policy." *Journal of Monetary Economics* 12, no. 1: 101–21.

Benigno, Pierpaolo, and Michael Woodford. 2005. "Inflation Stabilization and Welfare: The Case of a Distorted Steady State." *Journal of the European Economic Association* 3, no. 6: 1185–236.

Berger, Helge, Jakob De Haan, and Sylvester C. W. Eijffinger. 2001. "Central Bank Independence: An Update of Theory and Evidence." *Journal of Economic Surveys* 15, no. 1: 3–40.

Bernanke, Ben S., Thomas Laubach, Frederic S. Mishkin, and Adam S. Posen. 1999. *Inflation Targeting: Lessons from the International Experience*. Princeton, N.J.: Princeton University Press.

Bernanke, Ben S., and Frederic S. Mishkin. 1997. "Inflation Targeting: A New Framework for Monetary Policy?" *Journal of Economic Perspectives* 11, no. 2: 97–116.

Bianchi, Francesco. 2013. "Regime Switches, Agents' Beliefs, and Post–World War II U.S. Macroeconomic Dynamics." *Review of Economic Studies* 80, no. 2: 463–90.

Binder, Carola Conces. 2021. "Political Pressure on Central Banks." *Journal of Money, Credit, and Banking* 53, no. 4: 715–44.

Bloom, Nicholas, Mirko Draca, and John Van Reenen. 2016. "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity." *Review of Economic Studies* 83, no. 1: 87–117.

Bordo, Michael D., and Pierre Siklos. 2021. "The Transformation and Performance of Emerging Market Economies across the Great Divide of the Global Financial Crisis." In *Independence, Credibility, and Communication of Central Banking*, edited by Ernesto Pastén and Ricardo Reis. Santiago: Central Bank of Chile.

Buera, Francisco J., Alexander Monge-Naranjo, and Giorgio E. Primiceri. 2011. "Learning the Wealth of Nations." *Econometrica* 79, no. 1: 1–45.

Caldara, Dario, Sarah Conlisk, Matteo Iacoviello, and Maddie Penn. 2023. "Do Geopolitical Risks Raise or Lower Inflation?" Working Paper. https://www.matteoiacoviello.com/research_files/GPR_INFLATION_PAPER.pdf.

Caldara, Dario, and Matteo Iacoviello. 2022. "Measuring Geopolitical Risk." *American Economic Review* 112, no. 4: 1194–225.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12, no. 3: 383–98.

Campillo, Marta, and Jeffrey A. Miron. 1997. "Why Does Inflation Differ across Countries?" In *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer. Chicago: University of Chicago Press.

Cavallo, Alberto, Francesco Lippi, and Ken Miyahara. 2023. "Inflation and Misallocation in New Keynesian Models." Paper prepared for the ECB Forum on Central Banking, June 26–28, Sintra, Portugal. https://www.ecb.europa.eu/press/conferences/ecbforum/shared/pdf/2023/Lippi_paper.pdf.

Chari, Anusha, Peter Blair Henry, and Hector Reyes. 2021. "The Baker Hypothesis: Stabilization, Structural Reform, and Economic Growth." *Journal of Economic Perspectives* 35, no. 3: 83–108.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2011. "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review* 101, no. 3: 471–75.

Christiano, Lawrence. 2015. "Discussion of: 'Networks and the Macroeconomy: An Empirical Exploration' by Daron Acemoglu, Ufuk Akcigit, and William Kerr." *NBER Macroeconomics Annual* 30, no. 1: 346–73.

Clarida, Richard, Jordi Galí, and Mark Gertler. 1999. "The Science of Monetary Policy: A New Keynesian Perspective." *Journal of Economic Literature* 37, no. 4: 1661–707.

Cochrane, John H. 2023. *The Fiscal Theory of the Price Level.* Princeton, N.J.: Princeton University Press.

Coibion, Olivier, Yuriy Gorodnichenko, and Johannes F. Wieland. 2012. "The Optimal Inflation Rate in New Keynesian Models: Should Central Banks Raise Their Inflation Targets in Light of the Zero Lower Bound?" *Review of Economic Studies* 79, no. 4: 1371–406.

Congressional Budget Office (CBO). 2023. *The 2023 Long-Term Budget Outlook.* Washington: Author. https://www.cbo.gov/publication/59014.

Cukierman, Alex, Steven B. Webb, and Bilin Neyapti. 1992. "Measuring the Independence of Central Banks and Its Effect on Policy Outcomes." *World Bank Economic Review* 6, no. 3: 353–98.

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/medialibrary/media/research/staff_reports/sr1053.pdf.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics* 135, no. 2: 561–644.

De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. 2016. "Prices, Markups, and Trade Reform." *Econometrica* 84, no. 2: 445–510.

Dincer, Nergiz, Barry Eichengreen, and Petra Geraats. 2022. "Trends in Monetary Policy Transparency: Further Updates." *International Journal of Central Banking* 18, no. 1: 331–48.

Drechsel, Thomas. 2024. "Estimating the Effects of Political Pressure on the Fed: A Narrative Approach with New Data." Working Paper. https://econweb.umd.edu/~drechsel/papers/drechsel_political_pressure_shocks.pdf.

Dynan, Karen. 2023. "High and Rising US Federal Debt: Causes and Implications." In *Building a More Resilient US Economy*, edited by Melissa S. Kearney, Justin Schardin, and Luke Pardue. Washington: Aspen Institute.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2015. "Competition, Markups, and the Gains from International Trade." *American Economic Review* 105, no. 10: 3183–221.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2013. "The Decline of the U.S. Labor Share." *Brookings Papers on Economic Activity*, Fall: 1–63.

European Central Bank. 2023. *Eurosystem Staff Macroeconomic Projections for the Euro Area, December 2023*. Frankfurt: European Central Bank. https://www.ecb.europa.eu/pub/pdf/other/ecb.projections202312_eurosystemstaff~9a39ab5088.en.pdf.

Friedman, Milton. 1968. "The Role of Monetary Policy." *American Economic Review* 58, no. 1: 1–17.

Galí, Jordi. 2015. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Princeton, N.J.: Princeton University Press.

Goldberg, Pinelopi K., and Tristan Reed. 2023. "Is the Economy Deglobalizing? If So, Why? And What Is Next?" *Brookings Papers on Economic Activity,* Spring: 347–95.

Goodhart, Charles, and Manoj Pradhan. 2020. *The Great Demographic Reversal: Aging Societies, Waning Inequality, and an Inflation Revival*. London: Palgrave Macmillan.

Gopinath, Gita. 2023. "Europe in a Fragmented World." First Deputy Managing Director Remarks for the Bernhard Harms Prize, Kiel Institute for the World Economy, November 30. https://www.imf.org/en/News/Articles/2023/11/30/sp-fdmd-remarks-bernhard-harms-prize.

Guazzini, Vicente, Anastasia Leskova, and Massimo Meloni. 2023. "The Evolution of FDI Screening Mechanisms: Key Trends and Features." *Investment Policy Monitor*, Issue 25. New York: UNCTAD. https://unctad.org/system/files/official-document/diaepcbinf2023d2_en.pdf.

Ha, Jongrim, M. Ayhan Kose, and Franziska Ohnsorge. 2023. "One-Stop Source: A Global Database of Inflation." *Journal of International Monetary and Finance* 137: 102896.

Halac, Marina, and Pierre Yared. 2020. "Inflation Targeting under Political Pressure." In *Independence, Credibility, and Communication of Central Banking*, edited by Ernesto Pasten and Ricardo Reis. Santiago: Central Bank of Chile.

Halac, Marina, and Pierre Yared. 2022. "Instrument-Based versus Target-Based Rules." *Review of Economic Studies* 89, no. 1: 312–45.

International Monetary Fund. 2020. *Annual Report on Exchange Arrangements and Exchange Restrictions 2019.* Washington: Author. https://www.imf.org/en/Publications/Annual-Report-on-Exchange-Arrangements-and-Exchange-Restrictions/Issues/2020/08/10/Annual-Report-on-Exchange-Arrangements-and-Exchange-Restrictions-2019-47102.

International Monetary Fund. 2023a. "The High Cost of Global Economic Fragmentation." IMF Blog, August 28. https://www.imf.org/en/Blogs/Articles/2023/08/28/the-high-cost-of-global-economic-fragmentation.

International Monetary Fund. 2023b. *Climate Crossroads: Fiscal Policies in a Warming World.* Washington: Author. https://www.imf.org/en/Publications/FM/Issues/2023/10/10/fiscal-monitor-october-2023.

Jordà, Òscar, Sanjay R. Singh, and Alan M. Taylor. 2020. "The Long-Run Effects of Monetary Policy." Working Paper 26666. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26666.

Juhász, Réka, Nathan Lane, Emily Oehlsen, and Verónica C. Pérez. 2023. "The Who, What, When, and How of Industrial Policy: A Text-based Approach." Working Paper. SocArXiv. Ithaca, N.Y.: Cornell University. https://doi.org/10.31235/osf.io/uyxh9.

Juhász, Réka, Nathan J. Lane, and Dani Rodrik. 2023. "The New Economics of Industrial Policy." Working Paper 31538. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w31538.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline in the Labor Share." *Quarterly Journal of Economics* 129, no. 1: 61–103.

Karabarbounis, Loukas, and Brent Neiman. 2018. "Accounting for Factorless Income." *NBER Macroeconomics Annual* 33: 167–228.

Klebnikov, Sergei. 2023. "Cathie Wood Predicts AI Will Lead to Deflation and Drive Markets Higher." *Forbes*, October 4. https://www.forbes.com/sites/sergeiklebnikov/2023/10/04/cathie-wood-predicts-ai-will-lead-to-deflation-and-drive-markets-higher/?sh=18d2c3aac35f.

Ma, Yueran, and Kaspar Zimmermann. 2023. "Monetary Policy and Innovation." Working Paper 31698. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w31698.

Nakamura, Emi, and Jón Steinsson. 2008. "Five Facts about Prices: A Reevaluation of Menu Cost Models." *Quarterly Journal of Economics* 123, no. 4: 1415–64.

Nakamura, Emi, Jón Steinsson, Patrick Sun, and Daniel Villar. 2018. "The Elusive Costs of Inflation: Price Dispersion during the U.S. Great Inflation." *Quarterly Journal of Economics* 133, no. 4: 1933–80.

Nelson, Edward. 2022. "How Did It Happen? The Great Inflation of the 1970s and Lessons for Today." Finance and Economic Discussion Series 2022-37. Washington: Board of Governors of the Federal Reserve System. https://doi.org/10.17016/FEDS.2022.037.

Philippon, Thomas. 2019. *The Great Reversal: How America Gave Up on Free Markets*. Cambridge, Mass: Harvard University Press.

Primiceri, Giorgio E. 2006. "Why Inflation Rose and Fell: Policymakers' Beliefs and US Postwar Stabilization Policy." *Quarterly Journal of Economics* 121, no. 3: 867–901.

Qiu, Han, Hyun Song Shin, and Leanne Si Ying Zhang. 2023. "Mapping the Realignment of Global Value Chains." *Bank for International Settlements Bulletin* 78. https://www.bis.org/publ/bisbull78.htm.

Rogoff, Kenneth. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics* 100, no. 4: 1169–89.

Rogoff, Kenneth. 2003. "Globalization and Global Disinflation." *Federal Reserve Bank of Kansas City Economic Review* 88, no. 4: 45–78.

Rogoff, Kenneth. 2023. "The Stunning Resilience of Emerging Markets." Project Syndicate, October 31. https://www.project-syndicate.org/commentary/macroeconomic-orthodoxy-saved-emerging-markets-from-debt-crisis-by-kenneth-rogoff-2023-10.

Rogoff, Kenneth, Barbara Rossi, and Paul Schmelzing. 2022. "Long-Run Trends in Long-Maturity Real Rates 1311–2021." Working Paper 30475. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30475.

Romelli, Davide. 2022. "The Political Economy of Reforms in Central Bank Design: Evidence from a New Dataset." *Economic Policy* 37, no. 112: 641–88.

Sargent, T. J. 2001. *The Conquest of American Inflation*. Princeton, N.J.: Princeton University Press.

Sbordone, Argia. 2010. "Globalization and Inflation Dynamics: The Impact of Increased Competition." In *International Dimensions of Monetary Policy*, edited by Jordi Galí and Mark Gertler. Chicago: University of Chicago Press.

Schmitt-Grohé, Stephanie, and Martin Uribe. 2011. "The Optimal Rate of Inflation." In *Handbook of Monetary Economics, Volume 3B*, edited by Benjamin M. Friedman and Michael Woodford. Amsterdam: North-Holland.

Schreger, Jesse, Pierre Yared, and Emilio Zaratiegui. 2023. "Central Bank Credibility and Fiscal Responsibility." Working Paper 31246. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w31246.

Stansbury, Anna, and Lawrence H. Summers. 2020. "The Declining Worker Power Hypothesis: An Explanation for the Recent Evolution of the American Economy." *Brookings Papers on Economic Activity*, Spring: 1–77.

Stantcheva, Stefanie. 2024. “Why Do We Dislike Inflation?” *Brookings Papers on Economic Activity,* Spring, Conference Draft. (Final version included in this *BPEA* volume). https://www.brookings.edu/articles/why-do-we-dislike-inflation/.

Tobin, James. 1965. “Money and Economic Growth.” *Econometrica* 33, no. 4: 671–84.

US Federal Open Market Committee. 2023. *Summary of Economic Projections, December 2023.* Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/monetarypolicy/files/fomcprojtabl20231213.pdf.

Weise, Charles L. 2012. “Political Pressure on Monetary Policy during the US Great Inflation.” *American Economic Journal: Macroeconomics* 4, no. 2: 33–64.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, N.J.: Princeton University Press.

Yared, Pierre. 2019. “Rising Government Debt: Causes and Solutions for a Decades-Old Trend.” *Journal of Economic Perspectives* 33, no. 2: 115–40.

Comments and Discussion

COMMENT BY

DONALD KOHN Afrouzi, Halac, Rogoff, and Yared have raised a critical issue for economies and central banks. They argue that the global disinflation of the 1980s through the 2010s was greatly aided by a number of favorable developments—such as the sharp rise in trade and supply chain optimization as Eastern Europe and China joined the global trading system—that increased competition and lowered costs, making it easier for central banks and governments to adopt and achieve inflation targets. But these developments will not be repeated going forward, and some have shown signs of going into reverse; the resulting rise in costs and prices implies a less favorable trade-off of disinflation with output and employment. At the same time government debt levels have risen substantially relative to GDP after falling in earlier decades, pressuring interest rates higher and making those rates more salient for government budgeting. Those trends could, in turn, reawaken the "latent inflationary bias" in political systems and put pressure on central banks to be tolerant of higher inflation than would be optimal. The authors advocate for steps to increase central bank independence in order to resist these pressures. They embed their analysis and illustrate their concerns in a modified New Keynesian model in which, unlike in the standard model, aggregate supply and demand are nonlinear and the choice of a long-run inflation target affects real output.

In my comments, I won't be giving a detailed evaluation of the model, but I will note the difficulties I had in relating it to my lived experience as a monetary policymaker. Nonetheless, I will note that I agree with the authors that the evolving macroeconomic landscape, including sharp increases in government debt relative to income as well as shifts in globalization, could

Brookings Papers on Economic Activity, Spring 2024: 242–264

well escalate political pressures on central bank price stability mandates. I'm highly doubtful that we could see formal legislative action to strengthen central bank independence. Still, the Federal Reserve is undertaking a five-year review of its monetary policy framework—billed as a rethink of strategy, tools, and communications—and in the context of that review I believe there are steps the Federal Reserve could take to strengthen its commitment to price stability and enhance the public's understanding of the importance of that leg of its dual mandate.

RELATING THE MODEL TO POLICYMAKING The model results, and the slopes to long-run aggregate demand and supply that give a permanent trade-off of inflation and output, hinge on the distortions from imperfectly competitive firms facing constraints on how often they can adjust prices. The central bank can reduce the degree of distortions by aiming at higher inflation, which reduces firm share of output and increases labor share but also increases price dispersion, reducing demand. A more hawkish central bank lowering its inflation target will increase distortions, lowering labor share and demand while shifting aggregate supply outward through lower real wages; the effect of lower inflation on output is ambiguous.

In practice, monetary policymaking is focused on cyclical issues—how does policy need to be adjusted to achieve price stability and, for the Federal Reserve, its other legislative goal of maximum employment. Setting the inflation target has not considered the interaction of that target with the degree of monopolistic distortions or the associated labor share of output. For central banks, and I suspect for finance ministries where they are involved in establishing the inflation target, monopolistic distortions are the responsibility of the competition authorities, not the central bank and its inflation target.

Moreover, price stability has been seen as encouraging maximum output and employment over time (abstracting from issues of the effective lower bound on interest rates)—implying no long-run trade-off. Inflation distorts market signals and makes them hard to interpret so price stability unambiguously promotes efficiency. Recall Alan Greenspan's (2001) definition of price stability as inflation low enough that households and businesses don't need to take it into account when making decisions. And anchoring long-term expectations at the price stability target gives the central bank scope to lean against shortfalls in output without risking price stability. To be sure, labor share and the effects of imperfect competition do come into monetary policy discussions, but as factors affecting the dynamics of the path to achieving or maintaining price stability, not as factors influencing the level of the final target or its effect on output.

THE CHANGING ECONOMIC LANDSCAPE AND POLITICAL PRESSURES ON PRICE STABILITY MANDATES Global inflation dropped sharply from the early 1980s until 2000. The authors acknowledge the role in this development of increasing focus on price stability by central banks, the onset of explicit numerical inflation targets in the 1990s, and the reforms of governance structures to give central banks a degree of independence from short-term political pressures to pursue those mandates.

But they also point out that a number of developments over this period smoothed the path for disinflation, which helped bolster political support for the transition toward price stability. Some of these can be thought of as favorable supply shocks that lowered costs and prices without requiring any softening of demand and output. Globalization fits into this category as trade rose dramatically, responding to the sharp reduction in costs from containerization and decreases in tariffs and other trade barriers. That development effectively increased competition (for both firms and workers) in the context of the authors' model, lowering inflation and boosting income. Competition was also enhanced by deregulation and privatization as the Washington Consensus took hold. In the United States, a technology-driven increase in productivity growth from the mid-1990s until 2005 contributed to favorable inflation-growth combinations for a time. On the demand side, declining government debt to income—in the United States a rare run of federal surpluses in the late 1990s—reduced political pressure to keep funding costs down.

I agree with the authors that, at a minimum, these favorable shocks are not going to be repeated—for example, global trade volumes have leveled out since 2008—and some look like they are going into reverse. Tariffs and *friendshoring*—industrial policies to discourage imports of certain goods and encourage domestic production—will raise costs and increase domestic investment, boosting both inflation and equilibrium real interest rates. Rates will be further pressured higher by large persistent government deficits to increase defense spending in a geopolitically risky world, fund subsidies related to decarbonization, and serve the growing needs of an aging population, crowding out some private investment.

It's not clear how important these forces will be. Some of the most fundamental influences depressing equilibrium interest rates over recent decades—an aging population and modest productivity growth (pending a verdict on the effects of AI)—have not shifted.[1] Although globalization stopped increasing in 2008, the subsequent years until 2021 were marked

1. See International Monetary Fund (2023).

by very low inflation and real interest rates at or below zero. Some of the cost increases might be best thought of as onetime price level adjustments that are unlikely to result in higher inflation so long as longer-term expectations are anchored. And the "latent inflationary bias" of politicians should be mitigated to some extent by the public's intense dislike of inflation (Stantcheva 2024).

Still, financial markets participants have marked up their estimates of future r^* to 2+ in real terms and 4+ nominally. Cost pressures imply that trade-offs are not likely to be as favorable as before, possibly raising the unemployment rate consistent with low stable inflation. Higher interest costs will add to burgeoning fiscal deficits, to the difficulty of stabilizing debt to income as r rises relative to g, and to political discomfort. And one presidential candidate in 2024 has demonstrated in the past a predilection for trying to influence monetary policy decisions. Pressures on central banks, very much including the Federal Reserve, to hold down interest rates and tolerate greater inflation could well be more intense than they have been for several decades.

STRENGTHENING THE COMMITMENT TO AND PUBLIC UNDERSTANDING OF THE PRICE STABILITY MANDATE IN THE UNITED STATES: THE OPPORTUNITY OF THE FRAMEWORK REVIEW In light of these potential pressures, the authors recommend strengthening central bank independence to help central bankers continue to pursue price stability. Legislation to this end is highly unlikely in the United States, but the Federal Reserve has an opportunity to strengthen its commitment to price stability and reinforce public understanding of why that's important. That opportunity is the review of its monetary policy framework it first undertook in 2019–2020 and has announced it would repeat every five years, so in 2024–2025 for this round.

The annual "Statement on Longer-Run Goals and Monetary Policy Strategy" of the Federal Open Market Committee (FOMC) states that this review is to encompass policy strategy, tools, and communication, though the focus in 2019–2020 was on strategy.[2] A statement on goals and strategy was first adopted in 2012. In 2020, it was modified to better deal with the experience of the 2010s, a period of low inflation, often below the 2 percent target, and low interest rates, including considerable time when the FOMC's ability to cut the target federal funds rate to raise inflation to 2 percent had been constrained by the zero lower bound (ZLB). Periods at the ZLB threatened to cause the Federal Reserve to miss both its inflation and employment

2. The statement as adopted in 2020 and carried forward through January 2024 can be found at https://www.federalreserve.gov/newsevents/pressreleases/monetary20240131b.htm.

targets to the downside over time. The strategy statement published in 2020 therefore contained several pro-inflation asymmetries to offset the ZLB effect. First, monetary policy would seek inflation slightly above the target when it had been running below target for a while—with no mention of the response to a contingency of a run of above-target inflation; and second, policy would respond to shortfalls of employment from its estimate of sustainable maximum, but not to estimated overshoots unless inflation was already running above its target—there would be no preemption of rising inflation inferred from tight labor markets.[3]

In the event, of course, much of the period since the new framework was adopted and implemented in 2020 has been marked by inflation above the 2 percent target. The Federal Reserve reacted to the high inflation perhaps a bit late, but when it moved, it moved with speed and force. Although at this writing in spring of 2024, inflation is still notably above its target, it has fallen substantially from its peak, and expectations of inflation over the longer run appear to have been anchored around the target level, perhaps reflecting both the history of low inflation in previous decades and the evidence of policy determination to return inflation to its target. Nonetheless, in light of this more recent history and of the potential for escalating political pressures, the Federal Reserve should take the opportunity of its framework review to strengthen the public understanding of its commitment to price stability and make sure its strategy addresses periods of target overshooting as well as undershooting.

A good way to begin would be a thorough background examination of the experience since 2020. In retrospect, why was inflation so high and so poorly forecast? Did operating under the 2020 framework contribute to its level and persistence? What role might have been played by the forward guidance on interest rates and by the size and structure of asset purchases? What lessons can be learned from this history that might help shape the subjects and conclusions of the 2024–2025 framework review?

That study would seem to be a natural and essential starting point. A useful supplement would be a study of the forces highlighted by the authors that might raise price pressures and interest rates. How important are these pressures likely to be? Does the new 2025 framework need to be shaped in any particular way to address these possible developments, and if so, how?

The commitment to and understanding of the price stability target can be reinforced by consideration of how price stability should be defined.

3. An analysis of the framework and the forward guidance used to implement it can be found at Eggertsson and Kohn (2023).

In 2019, the Federal Reserve took its existing 2 percent target as given, explicitly ruling out an examination of whether that was the quantitative target that best fostered the public interest. A number of academics, concerned about low nominal rates constraining the response to negative demand shocks, have advocated for higher targets; the public would prefer lower—effectively zero.[4] The 2 percent target seems to meet the Greenspanian criteria referenced above, has a history in the United States as both an implicit and explicit definition of price stability, and is widely adopted internationally. But commitment and understanding would be strengthened by a careful examination and justification of the final choice—2 percent or otherwise—rather than leaving it as an arbitrary history-determined choice.

The commitment to price stability would be further strengthened by clarification of the maximum employment goal. The current framework notes that "the maximum level of employment is a broad-based and inclusive goal that is not directly measurable and changes over time owing largely to nonmonetary factors that affect the structure and dynamics of the labor market" (Federal Reserve Board of Governors 2020, par. 3). The phrase "broad-based and inclusive" was added in the 2020 revision and taken together with the asymmetrical approach to labor markets—paying attention to shortfalls but not overshoots—may have left the impression that the employment side of the dual mandate had been elevated relative to the price stability side. Individual FOMC participants have noted that maximum employment is the highest level of employment consistent with price stability, but that is not part of the long-run goals and strategy statement.[5] Including it would reinforce the consistency of the two goals and clarify that the Federal Reserve is not shaping its policy to correct for the distortions of imperfectly competitive firms, as in the authors' paper, or for historical inequities that have disadvantaged particular demographic or income groups.

The costs and benefits of the asymmetrical approach to maximum employment need an especially rigorous examination. The benefit is that it avoids policy firming that, in hindsight, unnecessarily constrains labor market expansion. But monetary policy acts on output and inflation with a lag. Because it reacts only to shortfalls of employment from maximum, the current framework strategy would rule out moving to a restrictive policy stance on the basis that labor markets were becoming tight enough to foster

4. On the higher target, see, for example, Blanchard (2022). For the public view, see Stantcheva (2024) on public disliking inflation included in this *BPEA* volume.

5. For example, Clarida (2022).

higher, above-target inflation down the road. That's a cost. Arguably, this asymmetry could have constrained the policy tightening of the mid to late 1980s and mid-1990s that were critical to consolidating and extending the gains against the inflation of 1979–1982 and ultimately anchoring expectations around the FOMC's target.

The new strategy statement needs to be robust to a wide variety of circumstances. It should retain the ability to deal with periods of very low inflation and interest rates. But it also should address more fully than the current statement the strategy for dealing with actual or prospective substantial and persistent inflation overshoots. Stress testing the new strategy statement against an array of scenarios would give the FOMC insight into the dynamics of their strategy and should reassure the public that the Federal Reserve had thought about how it would achieve its dual mandate whatever the source and consequence of the unexpected developments that might hit the economy.

Finally, the framework review should encompass a review of the FOMC's tools, especially the unconventional tools used at the ZLB—asset purchases and forward guidance about asset purchases and the target interest rate. What lessons can be drawn from the use of these instruments in 2020–2022? How should they be deployed in the future to assure progress toward price stability as well as maximum employment? Such an open inquiry would reinforce the public's understanding of the Federal Reserve's commitment to price stability, whatever pressures might descend on it in the future.

REFERENCES FOR THE KOHN COMMENT

Blanchard, Olivier. 2022. "It Is Time to Revisit the 2% Inflation Target." *Financial Times,* November 29. https://www.ft.com/content/02c8a9ac-b71d-4cef-a6ff-cac120d25588.

Clarida, Richard H. 2022. "The Federal Reserve's New Framework: Context and Consequences." Finance and Economics Discussion Series 2022-001. Washington: Board of Governors of the Federal Reserve System. https://doi.org/10.17016/FEDS.2022.001.

Eggertsson, Gauti B., and Donald Kohn. 2023. "The Inflation Surge of the 2020s: The Role of Monetary Policy." Hutchins Center Working Paper 87. Washington: The Brookings Institution. https://www.brookings.edu/articles/the-inflation-surge-of-the-2020s-the-role-of-monetary-policy/.

Federal Reserve Board of Governors. 2020. "Statement on Longer-Run Goals and Monetary Policy Strategy." Washington: Author. https://www.federalreserve.gov/monetarypolicy/review-of-monetary-policy-strategy-tools-and-communications-statement-on-longer-run-goals-monetary-policy-strategy.htm.

Greenspan, Alan. 2001. "Transparency in Monetary Policy." Speech given at the Economic Policy Conference, Federal Reserve Bank of St. Louis, October 11. https://www.federalreserve.gov/boarddocs/speeches/2001/20011011/default.htm.

International Monetary Fund. 2023. *World Economic Outlook: A Rocky Recovery.* Washington: Author. https://www.imf.org/en/Publications/WEO/Issues/2023/04/11/world-economic-outlook-april-2023.

Stantcheva, Stefanie. 2024. "Why Do We Dislike Inflation?" *Brookings Papers on Economic Activity,* Spring, Conference Draft. (Final version included in this *BPEA* volume). https://www.brookings.edu/articles/why-do-we-dislike-inflation/.

COMMENT BY

SILVANA TENREYRO This is a timely paper, studying how economic and political economy factors can interact to exert inflationary pressures on the economy. The analysis is based on a stylized model of aggregate demand and supply. The model is augmented to reflect central bank preferences that might differ from those of households, as in Rogoff (1985). Using this framework, the paper seeks to illustrate how past economic trends (e.g., globalization and deunionization) have exerted downward inflationary pressures, facilitating the task of achieving central banks' inflation targets. The analysis leads to a stark warning that a reversal of those trends might pose important challenges to central banks in the future. The key conclusion is that for inflation to remain low and stable, it is vital to maintain, and indeed reinforce, central bank independence and have in place a credible public debt policy.

The paper addresses a hugely important topic for policymakers and academics. It elegantly combines insightful ideas with model and data, leading to a new model narrative that underscores the risks to inflation and to the current monetary policy framework.

My comments zoom in on some aspects of the paper in the hope of clarifying to the broader readership its contribution to the literature and its connection with the practice of central banking.

THE MODEL The paper develops a simple model of aggregate demand and supply to carry out a positive analysis of long-run inflation. How is this model different from models used in central banks? First and foremost, the model is designed to think about political economy pressures that central banks might face in response to changes in the environment; those political pressures are not part of central bank models (rightly so).[1] However—and

1. It would be odd if, given their remits and the current institutional setting, central banks were to use a model in which, in some future, the central bank itself aimed off its own objectives or accommodated political pressures.

this is the risk highlighted by the paper—political pressures, under certain environments, might affect the behavior of central banks or, stretching a bit the model, could eventually lead to changes in mandates and policy frameworks.

Conceptually, the model seeks to capture how long-run inflation can be affected by the interaction between economic factors (e.g., the degree of monopoly power in the economy) and central banks' preferences. In the stylized model, those preferences are represented by the size of the labor income share targeted by a central bank, with a higher targeted labor share representing more "dovishness." In practice, this specification can be mapped into the more familiar "weight" that central bankers (or, perhaps more broadly, the monetary policy framework as reflected in their mandates) place on inflation stabilization versus a secondary objective of output stabilization: the more weight a central bank puts on output stabilization (over inflation), the higher the degree of dovishness.[2]

The model in this paper thus sits on a different layer of macroeconomic policy design, one that considers political economy risks. As such, it is distinct in its scope and ambition from models used by central banks; the latter are used for positive analysis to predict macroeconomic outcomes, or for normative analysis to optimize outcomes (e.g., the inflation path), given their mandates, over a finite (short- to medium-term) time horizon. By design, central bank models would not forecast future changes in inflation generated by political pressures.

To be sure, central banks can and do of course incorporate changing economic trends (e.g., deglobalization, market power, or demographics) in their models. The Bank of England, for example, adjusted the potential productivity growth trend for the UK economy after the Brexit referendum as a result of the country's expected loss in openness; similarly, most central banks adjusted trend productivity growth after the financial crisis. But central banks' models, by design, do not feature changes in political pressures that might, as the paper argues, lead to changes in long-term inflation.

A second difference between this paper's model and the models used in central banks is its simplicity, which allows for a clear comparative static analysis of the steady state. While a strength for the long-term comparative statics, for the analysis of transitional dynamics, this simplicity might be a bit more costly. The paper's transitional dynamic analysis as

2. The labor share would map into lambda in, for example, Carney's (2017) lambda speech.

well as the interpretation of particular inflationary episodes (such as the recent surge in inflation) could benefit from incorporating some of the features present in richer central bank models. Among other features, those models (1) have more realistic lead-lag structures (with the aim of matching impulse responses in the data, including the fact that monetary policy affects the economy with a significant lag); (2) encompass a number of additional frictions (e.g., financial and labor market frictions, and in some versions, present bias or other forms of bounded rationality); and (3) allow for investment/capital and more realistic open-economy dimensions.

The main modeling contribution of the paper lies in the derivation of the long-run aggregate supply (LRAS) and demand (LRAD) curves, rather than the specific shorter-term or transitional dynamics.

MODELING CHOICES The paper makes two important and realistic assumptions.

Nonzero inflation in steady state. A first assumption is that inflation can be nonzero in the steady state. This is a welcome feature of the analysis, consistent with targets of 2 percent in most advanced economies (and higher in many emerging or developing economies).

The model captures a trade-off generated by inflation: on the one hand, higher inflation helps offset the distortion from monopolistic pricing, while on the other hand, it leads to inefficient price dispersion, which causes a misallocation of resources. In highlighting that trade-off, the paper connects to the literature on optimal inflation, going back to Tobin's (1972) notion of inflation as the "grease in the wheels": with downward nominal rigidities, some inflation could be beneficial in helping adjust real wages and relative prices.[3]

The paper emphasizes that the slope of the LRAS curve is positive. This is surprising: while the short-term trade-off between inflation and the output gap is intuitive, it is less evident how the trade-off can be sustained in the long run, as forward-looking agents adjust their expectations in response to central banks' actions. In New Keynesian models with rational agents and Calvo price setting, the long-run Phillips curve is vertical or near vertical to a first-order approximation around zero steady-state inflation.[4]

3. See also Adam and Weber (2023), Adam, Alexandrov, and Weber (2023), Coibion, Gorodnichenko, and Wieland (2012), and Guerrieri and others (2021, 2023). The model could potentially be extended in the future to carry out normative analysis on the policy framework, including the derivation of optimal targets.

4. It is vertical in the limit in which the discount factor goes to one, corresponding to the parameter ρ in this paper going to zero.

On closer inspection, however, the LRAS formulation in the paper is also vertical or nearly vertical, as I explain next. To see this, note that the LRAS relation is given by the equation:

$$\pi = \frac{\lambda\left(\rho + \lambda - \pi^*\right)}{\rho - \pi^*}\left[y + \frac{\varphi}{1+\varphi}\log\mu + \log\frac{\sigma\left(\tau+1\right)}{\sigma-1}\right],$$

where the LRAS slope is given by $\frac{\lambda\left(\rho + \lambda - \pi^*\right)}{\rho - \pi^*}$; λ denotes the frequency of price adjustment; ρ is the household discount factor; and π^* is the long-run value of inflation.

The formula allows for the possibility of an exactly vertical curve or infinite slope. It also permits a backward-bending Phillips curve. More generally, for reasonable numerical values, the resulting slope of the LRAS is very large in absolute values. Let us walk through some interesting special cases.

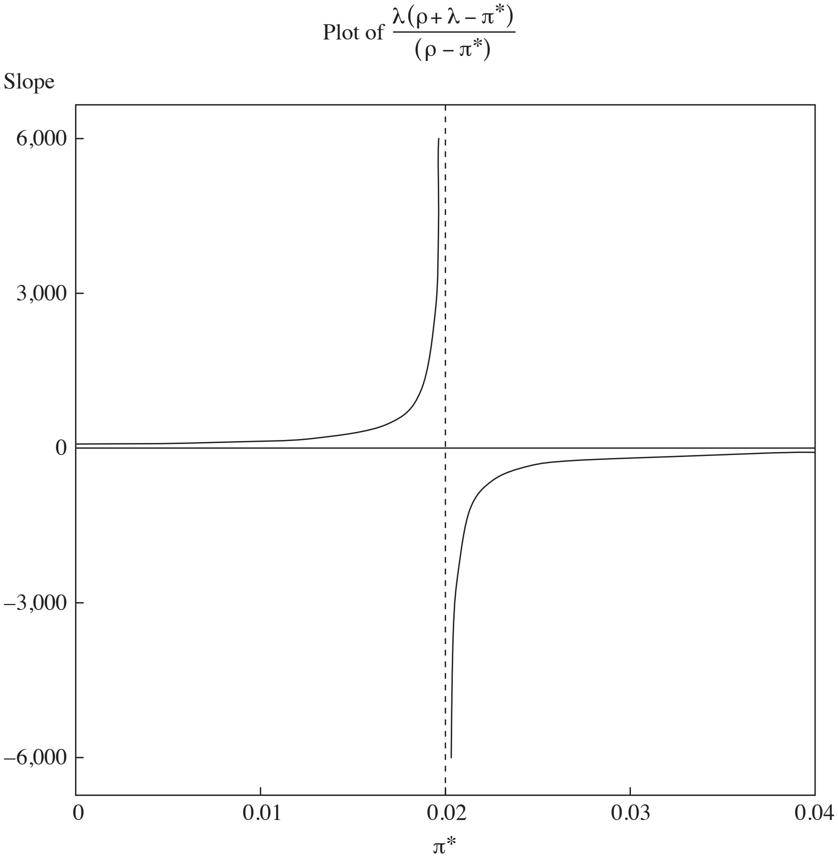
When $\pi^* = 0$, we have the more familiar expression for the LRAS or structural Phillips curve slope, $\frac{\lambda\left(\rho+\lambda\right)}{\rho}$, which converges to infinity as $\rho \to 0$. For a positive steady-state inflation, $\pi^* > 0$, the LRAS becomes vertical as $\rho \to \pi^*$. The LRAS slope turns negative when (1) $\rho < \pi^*$ and $\lambda + \rho - \pi^* > 0$ or (2) when $\rho > \pi^*$ and $\lambda + \rho - \pi^* < 0$.

More concretely, for a calibration of $\lambda = 1.2$ (as in the paper) and a discount rate of $\rho = 2\%$ (which seems reasonable), the LRAS slope becomes vertical at $\pi^* = 2\%$ and negative for $\pi^* > 2\%$, as illustrated in figure 1.

The figure, however, masks what happens away from $\pi^* = 2\%$. Even before becoming infinite, the values of the LRAS slope are also very high, as shown in table 1 for a range of selected π^* values and the same calibration of the other parameters as above. This implies that, in practice, the long-term trade-offs are not, in a quantitative sense, feasible, as the LRAS is practically inelastic. (Note that given that as the LRAS variables are expressed in log deviations from their steady states, the slope corresponds to the inverse of the LRAS elasticity, implying a near-zero long-run supply elasticity.)

Varying ρ changes the point at which the LRAS becomes exactly vertical, so the calibration of this parameter is important. However, as before, the slope of the LRAS is still very high in absolute values even when away from the asymptote. For example, for $\rho = 4\%$, which is the value preferred

Figure 1. Slope of the LRAS as a Function of Steady-State Inflation π^*



Source: Author's illustration.

Note: The figure shows the slope of the LRAS for $\lambda = 1.2$ and $\rho = 2\%$ as a function of the steady-state value of inflation, π^*. The slope passes to infinity at $\pi^* = 2\%$.

Table 1. Slope of the LRAS as a Function of Steady-State Inflation π^*

	LRAS slope	
π^* *(%)*	$\rho = 2\%$	$\rho = 4\%$
0.00	73.2	37.2
1.00	145.2	49.2
2.00	∞	73.2
3.00	–142.8	145.2
4.00	–70.8	∞
5.00	–46.8	–142.8
6.00	–34.8	–70.8

Source: Author's illustration.

Note: The table shows the slope of the LRAS curve for $\lambda = 1.2$ and $\rho = 2\%$ and $\rho = 4\%$ for selected values of steady-state inflation, π^*.

by the authors, the LRAS is decidedly inelastic even at lower values of π^*, as illustrated in table 1.

Perhaps it is fitting to address a misconception regarding New Keynesian models. In general, these models do not automatically generate a zero-inflation steady state: there is nothing in the model that ensures convergence to a zero-inflation (or 2 percent inflation) steady state; on the contrary, if the "wrong" policies are taken, inflation would end up above or below the 2 percent target in the long term.[5]

Though the paper deviates from the zero-inflation steady state, it follows closely other assumptions made in the simple New Keynesian model. In that setting, any price dispersion is inefficient, following the assumptions of symmetric preferences, concave utility over varieties, and similar technology (and common shocks) across varieties. In a richer setting with multiple sectors subject to different shocks and different degrees of price rigidities across sectors, the concept of price dispersion and its implication of efficiency is more nuanced. To be concrete, when an uneven shock (say, to gas prices) hits sectors differently (e.g., restaurants are far more affected than grocery stores), one might expect an increase in price dispersion, reflecting the uneven impact of the gas price shock. The change in price dispersion in this case can be efficient—it is the outcome of the price system doing its job. (An optimizing social planner would not want to fully stop those price signals, which facilitate the reallocation of resources in the face of shocks.) The pandemic and the energy price shocks are examples in which changes in relative prices (and dispersion) can be the efficient outcome (unlike in the simpler New Keynesian models); when combined with downward nominal rigidities, this can justify a temporary higher level of inflation.[6]

Lack of commitment. A second assumption in the paper is lack of commitment. The word *commitment* has different meanings among academics and practitioners. In the jargon of the academic literature, commitment means that the central bank decides at time zero a precise state-contingent policy path for the infinite set of future periods and states of the world. In the context of central banking, departing from the literature's definition of commitment is a realistic assumption, given that, in practice, central banks can only commit to their mandates and optimize outcomes over finite policy horizons. One could say that there is effectively discretion,

5. Another way to characterize this is that the model requires the specification of monetary policy behavior (the monetary policy rule) to be consistent with the desired long-run inflation rate. Put differently, it is the monetary policy rule that determines inflation in the long run.

6. See Guerrieri and others (2021, 2023).

or rather "limited commitment," over a rolling period of, say, three to five years. Why not longer? Because the current monetary policy board members cannot commit the decisions or votes of future board members.[7] A perhaps more fitting description is Bernanke's (2003) notion of "constrained discretion," which entails a middle ground between the academic extremes of full discretion and commitment. This notion still requires a commitment by central bankers, both through words and actions to price stability (however defined in their mandates).

In discussing commitment with a broader audience, it is hence important to emphasize the distinction between the meaning in the literature (commitment to an infinite state-contingent policy path) and the common understanding by market participants and other practitioners for which the term *commitment* is typically reserved for the mandate: are central banks committed to their mandates? This commitment to the mandate in practice is still consistent with the optimal "discretion" outcome in the literature, as long as central bankers have realistic expectations of the output potential of the economy—more on this later. Importantly, as pointed out by Giannoni (2020), the period-by-period optimization (or discretion) of a loss function (characterizing the mandate) leads to a strict Taylor-type rule (which practitioners outside academia might call "commitment").

CENTRAL BANK OBJECTIVES The paper assumes that the central banker in charge of policy seeks to optimize a social welfare function that considers all (possibly changing) distortions in the economy. In practice, central banks have much narrower mandates. Hence, a natural question is: can or do central banks aim off their narrow targets to improve social welfare?

Regarding feasibility, while it is true that objectives of full employment or output potential are not as precisely defined as inflation targets, there are two important lessons from central bank practice and theory, in particular from contributions of Barro and Gordon (1983) and Rogoff (1985), that

7. One could regard some announcements by central banks as trying to commit future members' policy actions. The key question is whether such announcements are credible, given that different decision makers may be in charge when the time comes to make good on the promise. There is an intermediate equilibrium concept of "loose commitment" (in which the policymaker operates under commitment but with a constant per period probability that previous commitments are abandoned). That may approximate central bank behavior somewhat better in certain cases. In a more complex model with endogenous state variables, the "discretionary" policymaker at date t realizes that their decisions can affect the state of the economy inherited by the date $t + 1$ policymaker and therefore takes this into account. Since the same logic holds for the policymaker at date $t + 1$, the discretionary policy problem becomes dynamic and intertemporal. However, the policymaker at t cannot directly control policy actions in future periods and can only influence those policies via the effects on the endogenous state variables.

can tackle the imprecision. The first is that central bank independence is a necessary condition for a sound conduct of monetary policy; the second is that central banks should target realistic estimates of the noninflationary (or inflation target–consistent) output potential. A central bank aiming for a higher level of activity than what would be consistent with inflation at target is bound to fail in fulfilling its inflation remit. This is well understood within the central bank community today.

In the simplest version of the New Keynesian model, it is typically assumed that the fiscal authority can correct the monopoly distortion with a labor subsidy, so that the flexible-price equilibrium level of output is efficient.[8] But realistically, absent the fiscal correction, central banks can only aim for the flexible-price equilibrium level of output, whether or not it is efficient. If a central bank aims to stimulate the economy beyond the inflation target–consistent level of output (trying to offset distortionary markups, for example), that will lead to an inflationary bias and a persistent deviation from target.[9]

The threat of an inflationary bias is the reason why there is a big effort in central banks to estimate the target-consistent output potential.[10] The inflationary bias is probably also why most central bank mandates give primacy to the inflation target over full employment, with some short-term flexibility in the face of temporary (supply) shocks.[11]

A different question is whether it pays for central bankers to deviate from their narrow targets and attempt to offset distortions, improving welfare. In advanced economies at least, deviations from targets today are costly for central bankers. Their performance is constantly scrutinized by media, parliamentary bodies, market participants, academics, and others. And there is a body of expertise ready to detect attempts at deviations.[12]

8. See Galí (2015) for a discussion of the efficient versus the distorted steady state.

9. And it is not obvious that the estimation errors should be one-sided (always estimating output potential above the true level); central banks can make mistakes, but over time, as the estimation model's performance is confronted with inflation outturns (and other outcomes), estimation and judgment would lead to convergence to the true values.

10. In the jargon of the literature, the target-consistent level of output corresponds to the flexible-price equilibrium level of output.

11. The logic to that short-term flexibility is that, given lags in transmission, monetary policy cannot offset the shock immediately (and if short-lived, the shock might disappear before policy has full effect).

12. Some would argue that it is much easier to detect and be penalized for missing the inflation target (vis-à-vis other objectives) since inflation is easier to measure than abstract concepts like the output gap or full employment. Given how much people dislike inflation, this would be a deterrent even to the most populist leaders; markets might also penalize such a move sooner or later, making it costly for politicians to attempt to change remits or institutional frameworks.

CHANGING ENVIRONMENT Of course, the main point of the paper is that the status quo could change. Political pressures may outweigh the pressure from public scrutiny and lead central bankers to aim for output above potential (in the model, a higher labor share) or a change in remits; or, perhaps, the changing environment might cause governments to remove or diminish central bank independence. This is the key question and challenge posed by the paper.

The paper is concerned specifically with changes in economic trends. It argues that globalization and the fall in union power made lives easier for central banks, effectively lessening the trade-offs between activity and inflation. In addition, lower indebtedness in the recent past (compared to now, and most notably among emerging economies) meant that there was less of an incentive to inflate away the debt.

I would also note that in the 1990s and early 2000s, there were no big negative supply shocks, a very different scenario from the 1970s and 1980s.[13] And certainly different from the early 2020s, which in a space of less than three years have witnessed a most remarkable concentration of rare events (particularly in Europe and the United Kingdom where the energy price increase alone, triggered by the Russian invasion of Ukraine, represented a shock comparable to, if not bigger than, the oil shock of the 1970s).

Despite this, central banks around the world have been focused on returning inflation to target. In the United Kingdom, consumption at the time of writing is 2 percent below what it was before the pandemic. In the euro area, consumption is just above its pre-COVID-19 level. The US economy is an exception, with consumption 11 percent above the pre-pandemic level, though still below pre-pandemic trends.[14] There is no sign that central banks in advanced economies, or indeed in many emerging economies and developing countries, have tried to push consumption or output higher.

It is important in the discussion to distinguish between changes in trends (that eventually can be foreseen) and unexpected (trade-off inducing) shocks.

13. While the financial crisis entailed a sharp loss in productivity, demand adjusted significantly, leading on net to a period of low inflation.

14. UK Office for National Statistics, "Household Final Consumption Expenditure: National Concept CVM SA–£m," https://www.ons.gov.uk/economy/nationalaccounts/satelliteaccounts/timeseries/abjr/pn2; US Bureau of Economic Analysis, "Real Personal Consumption Expenditures [PCECC96]," retrieved from FRED, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/PCECC96; Eurostat, "GDP and Main Components (Output, Expenditure, and Income)," https://ec.europa.eu/eurostat/databrowser/view/namq_10_gdp__custom_8299778/default/line?lang=en.

In the first case, central banks would need to change estimates of potential (as they eventually did post-financial crisis or post-Brexit); the question in the paper is: will they? As for unexpected shocks, if it is an isolated event, the orthodox response would be to accommodate in part, making sure that inflation returns to target; but if shocks become so frequent that they change the trend in potential output, we are back to the first case—and the same question posed by the paper.

I turn now to the question of changing trends and the impact on inflation.

GLOBALIZATION AND MARKUPS While the partial equilibrium effect of globalization might be intuitive, the general equilibrium effects are less obvious. A standard conceptualization of globalization, highlighted by Goodhart and Pradhan (2020), is that globalization lowered the prices of imported goods, and to the extent that the process was gradual, it led to lower imported goods price inflation. However, in general equilibrium, this improvement in terms of trade also increased real incomes, raising private demand and pushing up services inflation. The impact on inflation is not a priori obvious.[15] Deglobalization, conversely, should reduce real incomes and eventually demand, lowering domestic inflationary pressures. Indeed, globalization peaked in 2008, but we had not seen a reversal on inflationary pressures during the 2008–2019 period. On the contrary, inflation kept undershooting targets and central banks did not need to raise rates.

The paper conceptualizes deglobalization as an increase in the level of markups, as the economy becomes less competitive. This leads to a contraction in supply, an intuitive partial equilibrium effect. Going beyond the partial equilibrium effect, in practice, this redistribution away from workers may lead to a reduction in aggregate demand if profits accrue to agents with low marginal propensity to consume. It is not a priori obvious that the net effect of these forces would be inflationary.[16] But if, as in the model, the central bank tries to keep the labor share constant (equivalent to trying to stimulate the economy over the new, lower potential level of output), that would be inflationary. The point to stress is that it is not about inflationary pressures from the trends themselves, which could be muted in general equilibrium by private demand responses; it is instead a matter of lower

15. See Ambrosino and others (2024) who show the impact of deglobalization depends on how demand responds to lower real incomes caused by higher import prices.

16. Sbordone (2007) studies the link between globalization, markups, and inflation. She shows how key theoretical channels cancel out, leading to a muted impact on inflation; her theoretical result is matched by limited inflationary effects found in the numerous empirical studies she discusses.

output and real incomes, which might lead central banks (or governments) to push for more stimulus. Though the distinction might sound academic, the key challenge is the political pressure stemming from lower potential growth.

POLITICAL PRESSURES AND THE ROLE OF RESEARCH The risk in a context of low growth potential is that governments will put pressures on central banks to stimulate output.[17] (An alternative motivation, not developed in the model, but mentioned in the paper, is that the pressure to inflate comes because of higher levels of indebtedness.) The pressure could materialize in different forms. Governments might undermine or take away central bank independence; or they could persuade central banks to aim off their inflation targets to stimulate the economy or inflate away the debt.[18] Another manifestation of the pressure could be directly through a change in remit.

On the first possibility, there is probably near consensus among economists that undermining or taking away independence, or attempting to manipulate central banks would be a disastrous outcome. On the second option, there is a debate still unsettled on the optimal inflation target (Blanchard 2022); more generally, in a flexible inflation targeting regime, more debate is needed on how to stipulate the mandate in the face of unexpected supply shocks. This paper offers a useful model to frame that debate. In that context, there is an important role for academic and policy institutions (like the Brookings Institution) to play in this debate. After all, the academic literature (Barro and Gordon 1983; Rogoff 1985; Alesina and Summers 1993) was hugely influential in leading to central bank independence.

CONCLUDING REMARK Let me conclude by emphasizing that this is an important paper, underscoring a risk to central bank independence that we all need to take seriously. I hope the paper, and the risk it highlights, will be an important input in the exchange between academics and policymakers.

17. See Drechsel (2023) for an empirical study of political pressures on the Federal Reserve Bank.

18. It is far from obvious that with so much knowledge accumulated over the years, central banks themselves would try to systematically aim off the level of output consistent with inflation at the new chosen target. But if, hypothetically, a political appointee reveals with words or actions that there is a new output objective inconsistent with the stated inflation target, that would likely trigger sharp market reactions, which would be costly to the government (especially a highly indebted one). So it becomes important to think about the sequencing that will make turning dovish a politically appealing option. This is particularly relevant in the current context. After the recent inflation overshoot and people's dissatisfaction with high inflation, the political bias will turn to run in the opposite direction, that is, against inflation. Similarly, in line with Rotemberg's (2013) theory of central bank's "penitence," central banks will be more likely to err on the side of being too hawkish.

REFERENCES FOR THE TENREYRO COMMENT

Adam, Klaus, Andrey Alexandrov, and Henning Weber. 2023. “Inflation Distorts Relative Prices: Theory and Evidence.” Discussion Paper 18088. Paris: Centre for Economic Policy Research.

Adam, Klaus, and Henning Weber. 2023. “Estimating the Optimal Inflation Target from Trends in Relative Prices.” *American Economic Journal: Macroeconomics* 15, no. 3: 1–42.

Alesina, Alberto, and Lawrence H. Summers. 1993. “Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence.” *Journal of Money, Credit and Banking* 25, no. 2: 151–62.

Ambrosino, Ludovica, Jenny Chan, and Silvana Tenreyro. 2024. “Trade Fragmentation, Inflationary Pressures and Monetary Policy.” London School of Economics, manuscript.

Barro, Robert J., and David B. Gordon. 1983. “Rules, Discretion and Reputation in a Model of Monetary Policy.” *Journal of Monetary Economics* 12, no. 1: 101–21.

Bernanke, Ben S. 2003. “Constrained Discretion and Monetary Policy.” Remarks made at the Money Marketeers of New York University, New York, N.Y., February 3. https://www.federalreserve.gov/boarddocs/speeches/2003/20030203/.

Blanchard, Olivier. 2022. “It Is Time to Revisit the 2% Inflation Target.” *Financial Times*, November 28. https://www.ft.com/content/02c8a9ac-b71d-4cef-a6ff-cac120d25588.

Carney, Mark. 2017. “Lambda.” Speech given at the London School of Economics, January 16. https://www.bankofengland.co.uk/-/media/boe/files/speech/2017/lambda.pdf.

Coibion, Olivier, Yuriy Gorodnichenko, and Johannes F. Wieland. 2012. “The Optimal Inflation Rate in New Keynesian Models: Should Central Banks Raise Their Inflation Targets in Light of the Zero Lower Bound?” *Review of Economic Studies* 79, no. 4: 1371–406.

Drechsel, Thomas. 2023. “Estimating the Effects of Political Pressure on the Fed: A Narrative Approach with New Data.” Discussion Paper 18612. Paris: Centre for Economic Policy Research. https://cepr.org/publications/dp18612.

Galí, Jordi. 2015. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Princeton, N.J.: Princeton University Press.

Giannoni, Marc P. 2020. “Comment on ‘Optimal Inflation and the Identification of the Phillips Curve.’” *NBER Macroeconomics Annual* 34: 256–66.

Goodhart, Charles, and Manoj Pradhan. 2020. “The Great Demographic Reversal.” *Economic Affairs* 40, no. 3: 436–45.

Guerrieri, Veronica, Guido Lorenzoni, Ludwig Straub, and Iván Werning. 2021. “Monetary Policy in Times of Structural Reallocation.” *Economic Policy Symposium Proceedings: Macroeconomic Policy in an Uneven Economy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Guerrieri, Veronica, Michala Marcussen, Lucrezia Reichlin, and Silvana Tenreyro. 2023. *Geneva 26: The Art and Science of Patience: Relative Prices and Inflation*. Paris: Centre for Economic Policy Research.

Rogoff, Kenneth. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics* 100, no. 4: 1169–89.

Rotemberg, Julio J. 2013. "Shifts in US Federal Reserve Goals and Tactics for Monetary Policy: A Role for Penitence?" *Journal of Economic Perspectives* 27, no. 4: 65–86.

Sbordone, Argia M. 2007. "Globalization and Inflation Dynamics: The Impact of Increased Competition." Working Paper 13556. Cambridge, Mass.: National Bureau of Economic Research.

Tobin, James. 1972. "Inflation and Unemployment." *American Economic Review* 62, no. 1/2: 1–18.

GENERAL DISCUSSION Robert Hall noted that a prominent feature of the paper is the departure from central bank commitment and referenced a paper by Stanley Fischer that documents how governments often successfully commit to fiscal policy, even if that policy is not optimal.[1] Hall wondered if a similar idea could apply to monetary policy. Following up on Hall's point, Maurice Obstfeld observed that the paper implicitly introduces fiscal policy as a factor pressuring the central bank, but he would also like to see the results where monetary-fiscal interaction is modeled explicitly. Obstfeld added that, even in a model without commitment, the central bank objective function implies that the monetary authority will respond to fiscal pressures. In response, Pierre Yared pointed to his work with Jesse Schreger and Emilio Zaratiegui about fiscal-monetary interaction.[2] Yared summarized two channels: first, the central bank is pressured to devalue existing government debt; and second, the central bank is pressured to reduce the cost of issuing new debt.

Jón Steinsson argued that commitment was crucial in keeping inflation expectations well anchored and the sacrifice ratio low during the COVID-19 pandemic. Because commitment has proven so important, he praised the authors for presenting a model where perfect commitment is

1. Stanley Fischer, "Dynamic Inconsistency, Cooperation and the Benevolent Dissembling Government," *Journal of Economic Dynamics and Control* 2 (1980): 93–107.

2. Jesse Schreger, Pierre Yared, and Emilio Zaratiegui, "Central Bank Credibility and Fiscal Responsibility," working paper 31246 (Cambridge, Mass.: National Bureau of Economic Research, 2023). https://www.nber.org/papers/w31246.

not taken for granted. Steinsson said assumptions of perfect commitment, often implicit in policy rules, prevent most models from having anything meaningful to say about inflation expectations or long-run inflation, which this paper overcomes. He encouraged others to follow suit in building models where the monetary authority cannot perfectly commit. Steinsson and Kenneth Rogoff both warned against an "end-of-history" bias; central bank independence—and therefore central banks' ability to credibly commit to policy—is a relatively new phenomenon, and it is not guaranteed into the future.

Athanasios Orphanides agreed with other participants that central bankers struggle in practice to perfectly commit to policy but contemplated whether the model featured too much discretion. Orphanides emphasized that constraining discretion is a useful tool for preserving price stability and asked the authors how to relate that lesson to the paper's policy advice. In particular, Orphanides was curious what lessons the paper has for central bank policy frameworks. Rogoff clarified that the model is about the world, not just the United States or a single institutional framework. Hassan Afrouzi also noted it is unclear whether the Federal Reserve's maximum employment mandate assumes flexible prices or includes inflationary distortions like those in the paper. To provide better policy guidance, Afrouzi advocated for thinking through which distortions should be included or excluded when assessing the monetary policy framework.

Steven Davis observed that a useful next step in the literature could be using text-based approaches to parse exactly which forces central bankers are responding to and tailoring future models to those forces. Davis also encouraged building on the textual analysis in Charles Weise's 2012 paper following a similar remark from discussant Silvana Tenreyro.[3]

Obstfeld inquired whether assuming a constant frequency of price changes is reasonable in a model where trend inflation can change over time and asked how the model would behave if that parameter was endogenous. Afrouzi responded that, if they assumed higher inflation leads to more price flexibility, it would increase the sacrifice ratio and reinforce the idea that central banks should be wary of inflationary pressures.

Andrew Atkeson talked about how term structure models struggle to reconcile movements in long-term interest rates and related that literature to the paper. Atkeson discussed the paper by Sharon Kozicki and

3. Charles L. Weise, "Political Pressures on Monetary Policy during the US Great Inflation," *American Economic Journal: Macroeconomics* 4, no. 2 (2012): 33–64.

P. A. Tinsley,[4] which tried to explain long rates through movements in the Federal Reserve's long-run inflation target, and the paper by Stein and Hanson,[5] which explained long rates through changes in the term premia. Atkeson wondered if a political economy model, like the one presented by the authors, could explain movements in long rates or underlying inflation compensation. Conversely, Atkeson pointed out that rising long-term inflation compensation could provide a warning of falling central bank credibility.

Jonathan Pingle asked the authors how to think about the magnitude of possible effects in the context of advanced economy central banks. He noted that the paper places a heavy emphasis on the inflationary effects of deglobalization, although estimates of the disinflationary effects of globalization have been on the order of 0.2 points per year for a decade, so a similarly sized reversal may not change central bank behavior.

David Romer remarked that he found certain aspects of the model unintuitive. First, Romer questioned the upward-sloping long-run aggregate supply (LRAS) curve and how to interpret it in a model where monetary policy is neutral in the long run. Romer agreed with discussant Donald Kohn that most central bankers do not conceptualize the LRAS curve as upward sloping. Romer also questioned how the long-run aggregate demand (LRAD) curve works through price dispersion. He argued that the LRAD curve seemed to work through supply effects because the consumption bundle value of a given amount of labor is determined by the level of inflation. Romer suggested that the simpler model presented in a paper by Rogoff might suffice to capture much of the essence of the ideas in the present paper.[6] Rogoff observed that one crucial difference between this work and his 1985 paper is the transition dynamics: the new model allows significant but temporary overshoots and undershoots when shifting to a new steady state.

Afrouzi noted that the upward-sloping LRAS curve arises from how firms think about incorporating inflationary pressures into current prices. When inflation is higher, firms weigh how much they want to adjust current prices. As long as they don't fully adjust current prices, the LRAS curve will slope upward. Afrouzi agreed with Romer's consumption bundle intuition

4. Sharon Kozicki and P. A. Tinsley, "Shifting Endpoints in the Term Structure of Interest Rates," *Journal of Monetary Economics* 47, no. 3 (2001): 615–52.

5. Samuel G. Hanson and Jeremy C. Stein, "Monetary Policy and Long-Term Real Rates," *Journal of Financial Economics* 115, no. 3 (2015): 429–48.

6. Kenneth Rogoff, "The Optimal Degree of Commitment to an Intermediate Monetary Target," *Quarterly Journal of Economics* 100, no. 4 (1985): 1169–89.

of LRAD, but he clarified how it comes into the model. Afrouzi said that economic resources are fixed, so given the distribution of prices, households determine how much they want to demand. Only later does the model aggregate those choices with the supply side, determining output. How much output comes from those fixed resources is interpreted as productivity.

Obstfeld discussed the channels through which globalization or deglobalization could affect inflation. Obstfeld observed that, in the short run, most research emphasizes the role of import prices. Longer term, he highlighted the work of Charles Goodhart and Manoj Pradhan who emphasized the role of China, India, and the former Soviet Bloc entering the world economy and greatly increasing the effective labor force, putting downward pressure on wages and inflation.[7] Goodhart and Pradhan contend that these forces will reverse, which Obstfeld argued is a useful lens to view the inflationary pressures in this paper. He also pointed to a paper by Argia Sbordone as a useful reference for thinking about globalization in open models.[8]

7. Charles Goodhart and Manoj Pradhan, *The Great Demographic Reversal: Ageing Societies, Waning Inequality, and an Inflation Revival* (London: Palgrave Macmillan, 2020).

8. Argia M. Sbordone, "Globalization and Inflation Dynamics: The Impact of Increased Competition," in *International Dimensions of Monetary Policy,* ed. Jordi Galí and Mark J. Gertler (Chicago: University of Chicago Press, 2010): 547–79.

ANDREW J. FIELDHOUSE
Texas A&M University

CHRISTOFFER KOCH
Empirical Research Partners

DAVID MUNRO
Middlebury College

SEAN HOWARD
Independent

The Emergence of a Uniform Business Cycle in the United States: Evidence from New Claims-Based Unemployment Data

ABSTRACT Using newly digitized unemployment insurance claims data, we construct historical monthly unemployment series for US states going back to January 1947. We validate our series, showing that they are highly correlated with the Bureau of Labor Statistics' state-level unemployment data, which are only available since January 1976, and capture consistent business cycle dynamics. We use our claims-based unemployment rates to study the postwar evolution of labor market adjustments to local demand shocks and state unemployment fluctuations around national recessions. We document: (1) a trend decrease in the dispersion of relative employment growth and unemployment across states; (2) an attenuation of relative employment, unemployment, and population responses to state-specific demand shocks in recent decades; and (3) a convergence across states in both the speed and degree to which unemployment recovers after recessions. These trends show the emergence of a national business cycle experienced more uniformly across US states, particularly since the 1960s. We present evidence suggesting that a convergence in states' industrial composition helps explain why a more uniform business cycle emerged

Conflict of Interest Disclosure: David Munro received support from MiddData, a data science initiative at Middlebury College. Christoffer Koch was a former employee of the International Monetary Fund whose communications department had the right to review this work prior to publication but did not inform the findings. 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 2024: 265–319

when it did. And states' increasingly similar experience in recessions may help explain why interstate migration became a weaker adjustment mechanism in recent decades.

Macroeconomists are increasingly leveraging panel data sets and regional heterogeneity to identify economic relationships.[1] There is also an increasing awareness that the unemployment rate is one of the best indicators of economic slack, particularly for business cycle analysis (Romer and Romer 2019). Unfortunately, official unemployment rate data for US states are only available from the Bureau of Labor Statistics (BLS) starting in 1976, greatly hampering historical state-level analyses. For instance, the rich literature on state-level labor market recoveries, regional business cycles, and state coincident economic indexes has largely been limited to starting around 1976.[2]

In this paper we present a newly developed monthly unemployment data set for US states that greatly expands the time horizon for work with state-level panel data. Our novel unemployment series are constructed from a large data set of newly digitized monthly unemployment insurance (UI) claims, pieced together from various historical reports published by the Department of Labor (DOL) and Social Security Administration (SSA). Together with available monthly data on nonfarm payroll employment, we compute an alternative claims-based unemployment rate that can be consistently constructed for US states from January 1947 to the present day. We validate our new data set by showing that our claims-based unemployment rates are highly correlated with official measures of unemployment, both state and national, in available overlapping samples.[3] We also use our claims-based unemployment rates to identify postwar peaks and troughs in state and national business cycles and to document that our new measures

1. For instance, Nakamura and Steinsson (2014) and Chodorow-Reich (2019) exploit regional heterogeneity to identify cross-sectional fiscal multipliers, and Hazell and others (2022) exploit regional heterogeneity to study the slope of the Phillips curve. See Glandon and others (2023) for an overview of the recent shift in empirical macro toward panel data and micro data.

2. See, for example, Blanchard and Katz (1992), Crone and Clayton-Matthews (2005), Owyang, Piger, and Wall (2005), Brown (2017), Dao, Furceri, and Loungani (2017), and Tasci and Zevanove (2019).

3. We use "official measures of unemployment" to refer to data that are produced and presently made available by federal statistical agencies. We discuss historical data availability in online appendix A.1.

capture consistent business cycle patterns as official measures of unemployment, such as inflection points and amplitude dynamics.

We revisit the classic question of how labor markets respond to local demand shocks, using our longer historical sample of claims-based unemployment rates to study how these responses have evolved over the full postwar sample. Building on Blanchard and Katz (1992) and Dao, Furceri, and Loungani (2017), among others, we estimate relative employment, population, and unemployment responses to a relative industry mix Bartik (1991) instrument in a panel local projections-instrumental variable (LP-IV) framework over the full postwar sample and subsamples. Our analysis shows that the response of relative population, proxying for interstate migration, has diminished since the start of the Great Moderation. Similarly, we find that the responses of relative employment and unemployment have diminished and become less persistent in recent decades, contrary to the highly persistent responses of relative employment documented by Blanchard and Katz (1992) and Dao, Furceri, and Loungani (2017). Lastly, we document that larger Bartik shocks drive these responses, but these larger relative shocks are less frequent and smaller in magnitude in more recent decades, which helps explain the attenuation of these relative labor market responses.

In addition to studying local labor market responses to demand shocks, we also use our novel data set to study patterns between state and national business cycles to better understand their coevolution. We use our claims-based unemployment series to study the evolving pace and nature of labor market recoveries following all postwar US recessions. Our analysis of unemployment recoveries follows the recent work of Hall and Kudlyak (2020) but does so at the state level, which was previously precluded by data limitations. Hall and Kudlyak (2020) document that recoveries in the US unemployment rate have been quite stable since the early 1960s, but the pace of recovery has decelerated markedly since the recoveries from earlier postwar recessions.[4] We corroborate this stylized fact with our new data set and find that the faster, early postwar recoveries are associated with greater heterogeneity in recovery rates across states, whereas states tend to experience more uniform recovery rates in more recent, slower national recoveries. We show that this deceleration and convergence in states' recovery

4. Hall and Kudlyak (2020) find that, on average, the US unemployment rate falls by 0.1 log points—or one-tenth of the peak unemployment rate—per year after recovery begins, until this relatively stable recovery rate is upended by the next recession, consistent with "plucking models" of the business cycle (Dupraz, Nakamura, and Steinsson 2024).

rates is robust to indexing to state-specific business cycle troughs around national business cycle troughs. We also document a convergence across states in the degree to which unemployment recovers after recessions since the late 1970s.

The evidence from our historical claims-based unemployment rates points toward the emergence of a national business cycle experienced more uniformly across US states, particularly since the 1960s–1970s. We show that the industrial composition of states' economies became increasingly similar to one another—with much of this convergence transpiring in the 1940s–1960s—which helps explain why a more uniform business cycle emerged across states when it did. States' increasingly common experience in recessions and recoveries, in turn, may help explain why interstate migration became a weaker adjustment mechanism in recent decades.

I. Data Set Construction

In this section, we first overview the digitization and data cleaning process for historical state-level UI claims. We discuss the construction of novel claims-based unemployment series from these newly digitized data. To validate our data set, we analyze the relationship between our claims-based unemployment series and official unemployment measures during available overlapping samples. Lastly, we model and present an alternative "fitted" claims-based unemployment series, which some practitioners might find more appropriate for their purposes.

I.A. Digitizing Historical Unemployment Claims

Monthly state-level UI claims are presently available in digital form dating back to January 1971 from the DOL's website; see online appendix A.1. Using scanned versions of printed reports previously published by the DOL and SSA, we backdated the publicly available data by digitizing monthly data on initial claims (IC) and continued claims (CC) back to December 1946 for all fifty states and the District of Columbia.[5] The historical claims data originate from one of a series of periodical reports: *Employment Security Activities*, *The Labor Market and Employment Security*, *Unemployment Insurance Statistics*, and the *Unemployment Insurance Review*. We were able to access most of these primary sources via HathiTrust or Google Books and supplemented missing publications with interlibrary

5. The sample start is chosen so a three-month centered moving average of claims is available back to January 1947.

loan requests or scans from the DOL's internal library. We were almost always able to track down high-quality scans that were easily legible, but we used data on changes in claims to guide digitization when merited, and we always used reported data on national aggregates as a cross-check with the sum of state claims; see online appendix A.2 for details. In total, just over 36,000 monthly observations were digitized.

Newly digitized UI claims data were merged with the DOL's publicly available state-level IC and CC data for regular state programs only, to be consistent with the historical claims data. After merging the series, we seasonally adjusted the full backdated IC and CC series using the Census Bureau's Win X-13 seasonal adjustment program. We also used Win X-13 to run a series of outlier tests, which identified roughly 200 potential outliers from roughly 91,000 observations (newly digitized and existing data combined). We manually checked each potential outlier to assess whether it represented a legitimate change in claims (e.g., a surge in Louisiana following Hurricane Katrina) or a "fat thumb" data coding issue (e.g., an implausible spike in Missouri exceeding the state's population); see online appendix A.2 for details on data cleaning and seasonal adjustment. The monthly claims data we digitized are aggregated by the DOL from weekly claims data collected by state UI offices, and we first convert these monthly claims to average weekly claims; this approach mimics the DOL's conversion of weekly data to average weekly data for calculating insured unemployment in a given month.[6]

I.B. Claims-Based Unemployment Rates for US States

Using these UI claims data, we construct monthly claims-based unemployment rates for all fifty states and the District of Columbia. Our claims-based unemployment rate draws conceptually on both the official unemployment rate estimated by the BLS—the ratio of unemployed workers to the labor force—as well as the insured unemployment rate (IUR) produced by the DOL Employment and Training Administration (ETA)—the ratio of average weekly continued claims divided by covered employment, that is, workers eligible for state or federal unemployment programs. We use initial and continued claims as an alternate measure of unemployed workers (the subset receiving regular state benefits) and measure this as a

6. In keeping with the DOL data for average weekly insured unemployment in a given month, monthly data are weighted by the split number of five-day workweeks in the month. We calculate the weights as the sum of workdays in each given month divided by five days for the workweek, ignoring the distinction of holidays.

ratio to employed workers plus these UI claimants, a related proxy for the labor force influenced by data limitations. Specifically, our claims-based unemployment rate for state *i* in month *t* is computed as

$$UR_{i,t}^{Claims} = \frac{IC_{i,t} + CC_{i,t}}{NP_{i,t} + IC_{i,t} + CC_{i,t}} \tag{1}$$

where $IC_{i,t}$ and $CC_{i,t}$ are average weekly claims for the month and $NP_{i,t}$ is nonfarm payroll employment from the Current Employment Statistics (CES)—the only state-level employment series presently available at a monthly frequency back to 1947.[7] The seasonally adjusted claims data can be rather noisy, particularly for initial claims, so we smooth the $IC_{i,t}$ and $CC_{i,t}$ series using a three-month centered moving average in constructing equation (1). We analogously construct a claims-based unemployment rate for the United States, aggregating seasonally adjusted average weekly claims and nonfarm payroll employment for all fifty states and Washington, DC.

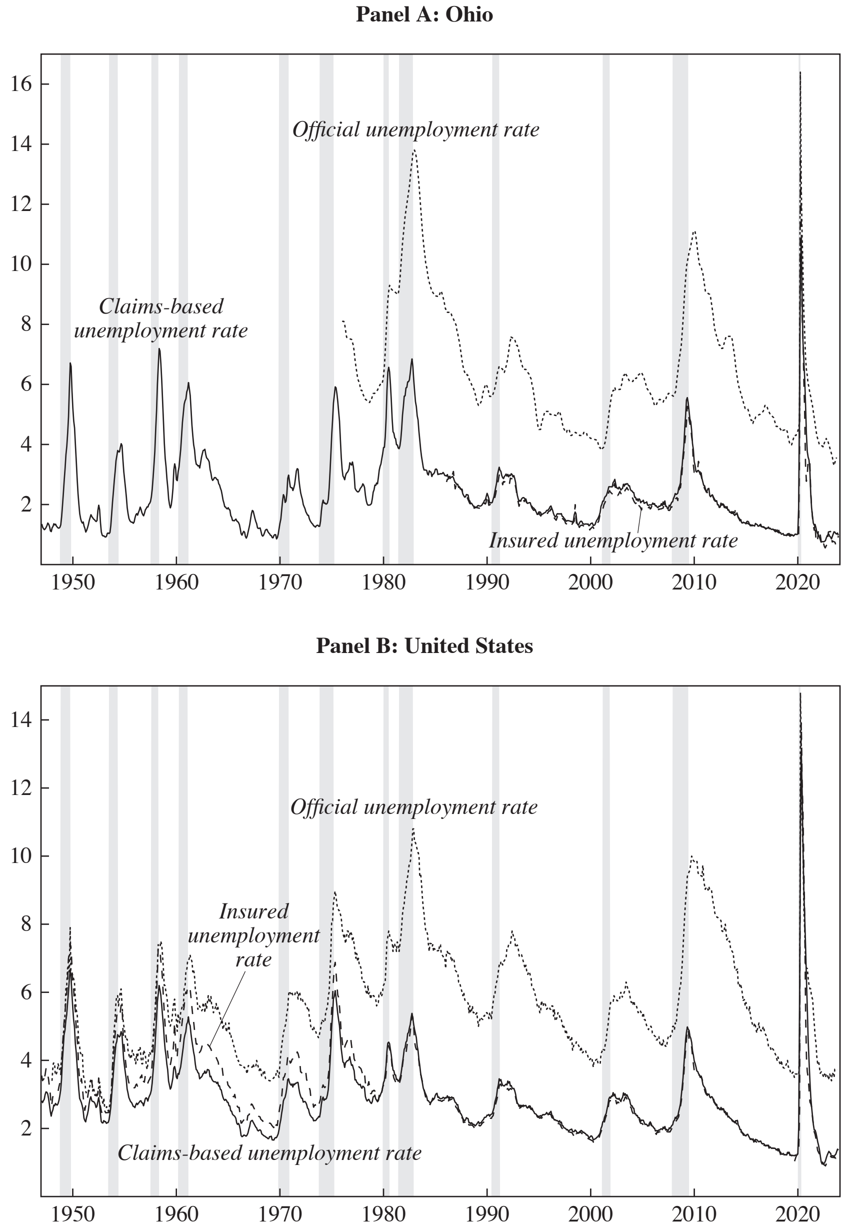
Panel A of figure 1 plots our claims-based unemployment rate (solid black line), the BLS unemployment rate (dotted gray line), and the DOL IUR (dashed black line) for Ohio, meant as an illustrative, representative large state; like all states, official data for Ohio start in January 1976 for the unemployment rate and February 1986 for the IUR.[8] The three unemployment series capture similar features of Ohio's business cycle in overlapping samples, such as identifying similarly timed local peaks and troughs. Figure 1, panel A, also underscores the practical benefit of our claims-based unemployment rates: relative to the official BLS data, our historical series offer nearly three additional decades of monthly state-level data, spanning six postwar national recessions as identified by the National Bureau of Economic Research (NBER) Business Cycle Dating Committee (gray bars).

Panel B of figure 1 plots our US claims-based unemployment rate (solid black line), along with the US unemployment rate (dotted gray) and the

7. Historical state-level data on covered employment are not consistently available at a monthly frequency from our primary sources. We seasonally adjust nonfarm payroll employment for each state using the Win X-13 program, as seasonally adjusted nonfarm payroll employment data for states are only available from the BLS starting in 1990. Several states do not have nonfarm employment data available since January 1947: data begin in January 1950 for Minnesota, in January 1956 for Michigan, in January 1958 for Hawaii, and in January 1960 for Alaska. Our claims-based unemployment rates are constrained to these later start dates for these four states.

8. The monthly IUR series is aggregated from weekly data that is not seasonally adjusted. We seasonally adjust the monthly IUR for Ohio using the Census Bureau's Win X-13 program.

Figure 1. Comparison of Official and Claims-Based Unemployment Rates for Ohio and the United States



Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS.

Note: Data sample from January 1947 to December 2023 (or when available). Gray bars denote NBER recession dates.

IUR (dashed gray), when available. The US unemployment rate starts in January 1948, and we digitized and backdated monthly US IUR data to January 1947, before official data begin in 1971; see online appendix A.3 for details. Figure 1, panel B, shows that our US claims-based unemployment rates are highly correlated with these official US measures over long overlapping samples and identify broadly consistent features of the aggregate business cycle. We discuss these relationships in more depth below.

I.C. Comparisons with Official Unemployment Measures

It must be emphasized that our claims-based unemployment rates measure labor market slack differently than either the official unemployment rate or the IUR.[9] Readers should not view our claims-based unemployment rates as an attempt to displace or backdate any other official measure; the motivation behind this new data set is simply to expand our ability to study the US economy and labor markets across states and over a longer time horizon. But as shown in this section, official measures and our claims-based measures of unemployment contain similar informational content about the degree and timing of labor market slack—the series are highly correlated at both the state and national level and identify comparable inflection points in economic activity.

The BLS calculates the US unemployment rate from the Current Population Survey (CPS), a monthly survey of roughly 60,000 households inquiring about their employment status; to be counted as unemployed, a worker must have been available for work but not been employed during the surveyed week and must have actively searched for work in the last four weeks or been expecting to be recalled following a temporary layoff. Official measures of state-level unemployment rates are intended to reflect the same definition of unemployment but are instead a statistical construct, derived in part from unemployment claims data.[10] And the IUR is calculated, for the United States as well as individual states, strictly from reported UI claims and coverage data.

It is immediately clear from figure 1 that our claims-based unemployment rates almost perfectly align with the IUR both in terms of levels and

9. See online appendix A.3 for further discussion of conceptual similarities, differences, and robustness checks.

10. The BLS Local Area Unemployment Statistics program uses data from the CES, CPS, and state UI programs to estimate state unemployment rates, but the methodology is something of a black box to the public; see the BLS Local Area Unemployment Statistics program web page.

inflection points; the correlation between Ohio's claims-based unemployment rate and the IUR is 0.98 in the overlapping sample. The key advantage of our state-level claims-based unemployment rates relative to IUR data is simply data availability: at a monthly frequency, digital state-level IUR data are only presently available from ETA back to February 1986, and data limitations appear to preclude digitizing and backdating consistent monthly state-level IUR series to the 1940s or 1950s; see online appendix A.1.

On the other hand, there is a level difference between our series and the official unemployment rates. More similar to the IUR, our claims-based unemployment rates are restricted to individuals qualifying for and claiming regular state UI benefits as reported weekly by state unemployment offices to the ETA or preceding agencies. This is a subset of the population surveyed by the CPS: state UI programs have typically excluded certain workers from benefit eligibility, notably agricultural and self-employed workers, while federal employees and veterans have usually been covered by separate federal UI programs.[11] Consequently, our claims-based unemployment rates should be strictly lower than the BLS unemployment rates because of the narrower pool of benefit-eligible workers and because anyone unemployed beyond the maximum duration for regular state UI benefits will drop out of our measure.

Figure 1, panel A, shows such an expected level difference between the official unemployment rate and our claims-based unemployment rate for Ohio, one that is quite stable. A stable level difference poses no impediment to business cycle analysis so long as the series are highly correlated (they are, with a correlation coefficient of 0.81) and identify comparable inflection points (they do); moreover, it could be differenced out or removed by detrending the series if desired.[12]

The level difference between the US series, however, shrinks moving back into the 1950s and 1940s, which is partly driven by our use of nonfarm payroll employment in the denominator of equation (1); when a larger share of workers are employed in agriculture and appear in the CPS survey measure of employment but not the CES measure we use, it mechanically pushes up our claims-based unemployment rate relative to the CPS

11. The larger pool of state and local government workers has been eligible for state UI programs for most of our sample and thus appear in our IC and CC measures. Official IUR series also typically focus on regular state UI programs and exclude federal UI programs, helping to explain the close match between the series seen in figure 1.

12. The correlation between the annual percentage point change in these two unemployment rates for Ohio is 0.82.

unemployment rate.[13] As would be expected, substituting the CPS measure of employment into equation (1) would hardly have any effect in recent decades but would gradually start pulling down our US claims-based unemployment rate moving back in time into the early postwar era, as seen in online appendix figure B.1. While using CPS employment would reduce the convergence in levels between the two US unemployment series in the 1940s–1960s, state-level CPS employment data are not available from the BLS until January 1976—precisely why we use the CES employment data.

Despite the time-varying level difference, our US claims-based unemployment rate identifies similar peaks and troughs as the BLS unemployment rate, as seen in figure 1, panel B, and quantified in section III.A, and the secular decline in the share of workers employed in agriculture would be absorbed by most detrending exercises. To illustrate this point, online appendix figure B.2 depicts the cyclical versus trend components of the official US unemployment rate (blue) and our US claims-based unemployment rate (red), both extracted using a Hodrick-Prescott filter (Hodrick and Prescott 1997). The detrended data underscore that the time-varying level difference between the series does not impede business cycle analysis: the inflection points between positive and negative cyclical unemployment line up nearly perfectly between the two detrended series, particularly so in the earlier decades when there was a greater divergence between total employment and nonfarm payroll employment. And the two filtered series are highly correlated, with a correlation coefficient of 0.89 for the full sample and 0.94 for the pre-1976 sample.[14]

One possible concern about our claims-based unemployment rate is that the maximum duration of benefits have, to a degree, changed over time; procyclical changes in benefit duration would be particularly problematic and would not be absorbed by detrending exercises. Our construction of claims-based unemployment rates from only regular state UI programs is partly intended to avoid such a confounding influence from standing or

13. The ratio of total farm employment to nonfarm payroll employment fell from 23.6 percent in 1947 to 6.4 percent in 1970 (Historical Statistics of the United States, Colonial Times to 1970, https://www.census.gov/library/publications/1975/compendia/hist_stats_colonial-1970.html); the CPS/CES employment ratio starts flattening out around 1970.

14. Hamilton (2018) raises compelling concerns about the Hodrick-Prescott filter and proposes an alternative linear forecasting method for detrending data, but the implied trend in unemployment is highly sensitive to any recent recession, quickly rising and thus generating rapid declines in cyclical unemployment. While we prefer the Hodrick-Prescott filter in this context, the Hamilton (2018) method generates similar inflection points and correlations in cyclical unemployment measures.

ad hoc benefit extensions during recessions. We also examine how the maximum benefit duration for regular state programs evolved using the State Unemployment Insurance Laws data set compiled by Maxim Massenkoff for 1970–2018, which we extend back to 1947 from scanned DOL reports.[15] Online appendix figure A.1 shows that the average maximum duration of benefits is quite stable throughout our sample of interest, and online appendix figure A.2 shows that the average duration of unemployment is almost always well below typical maximum benefit durations; as such, legislative changes to the maximum duration of regular benefits should have a minimal influence over time variation in our digitized UI claims.[16]

Another possible concern is that expansion of UI coverage in the early postwar era might be driving cyclical variation in our claims-based unemployment rate. Our use of nonfarm payroll employment in the denominator of equation (1) is always broader than the covered employment measure used in IUR calculations, but the share of workers covered by UI programs rose sharply in the early postwar era (McMurrer and Chasanov 1995), partly because of UI policy expansions and partly because of the shift from (mostly uncovered) farm labor to (mostly covered) nonfarm labor.

We digitize annual data on US covered employment for regular state programs back to 1945 to examine any concerning influence of expanding UI coverage for our claims-based unemployment rates. Online appendix figure A.4 shows that the ratio of US covered employment to nonfarm payroll employment was quite stable at roughly 72–75 percent from the 1940s through early 1970s, then—driven by two federal policy changes—jumps to roughly 95–97 percent by the late 1970s. Congruently, figure 1, panel B, shows that our US claims-based unemployment rate and our backdated IUR series line up almost seamlessly since the late 1970s, but diverge slightly in earlier years, before this coverage expansion. But the two series are consistently capturing the magnitude and timing of business

15. See Maxim Massenkoff, "State Unemployment Insurance Laws," http://maximmassenkoff.com/data.html, accessed October 25, 2021.

16. McMurrer and Chasanov (1995) similarly document stability in max benefit durations for regular state programs over much of our historical sample of study. As an additional robustness check we also compute an alternative variant of our claims-based unemployment rate only using IC data, which will not be affected by changes in maximum duration policies. Online appendix figure B.3 plots the (IC + CC) claims-based unemployment rate along with the IC-only variant; the two series track each other quite closely. This strong correlation highlights the fact that even after the trough of a business cycle, new separations from employment remain elevated for a significant period of time.

cycles throughout the sample, which is reassuring. Moreover, the last (and largest) UI coverage expansion occurs in the late 1970s, when the BLS state-level unemployment rates exist (these should be unaffected by coverage expansions); reassuringly, we do not observe a systematic change in our claims-based unemployment series relative to the official data around this period. But if desired, this federal expansion of UI coverage could be absorbed by detrending exercises; the Hodrick-Prescott-filtered cyclical components of the US claims-based unemployment rate and the IUR line up nearly identically (correlation of 0.98) throughout the sample and only the trend components diverge before the late 1970s; see online appendix figure B.2, panel B.[17]

Previewing some things that follow, much of our analysis below studies relative state-level variables that difference out national labor market averages, absorbing any common effect from federal policy changes. Moreover, the important changes we document regarding the convergence of unemployment dynamics across states and the emergence of a more uniform business cycle occur in the early postwar era, before the UI coverage expansions of the 1970s, meaning that these results are not being driven or biased by any spurious variation from those policy changes.

I.D. Fitted Claims-Based Unemployment Rates

Given the distinctions between the BLS state-level unemployment rates and our (unfitted) claims-based unemployment rates, we also estimate an alternative "fitted" measure of state unemployment rates using a statistical model of the relationship between the two series since January 1976. As the BLS uses UI claims as an input into their (not publicly known) statistical model, our fitting exercise explores how much informational content UI claims alone have for official state-level unemployment rates, since this is effectively the data world that exists pre-1976, before CPS micro data are available.[18] If a good fit to official unemployment rates is achieved with UI claims, that helps build confidence that the claims data capture consistent features of state-level labor markets, or even more mechanically, that claims are a key input to the BLS statistical model. The regression

17. As another related robustness check, we also compute an alternative annual version of our claims-based unemployment rates using state-level covered employment instead of nonfarm payroll employment; the Hodrick-Prescott-filtered series are nearly a perfect match with that of our claims-based unemployment rates for every state, underscoring that a similar degree of labor market slack is being driven by claims, not employment data; see online appendix figure B.4.

18. Similarly, more detailed CES data are not available until 1990.

framework we choose captures the idea that a state's unemployment rate is likely higher than the national rate when that state is experiencing a higher claims-based unemployment rate relative to the national claims-based rate; it also reflects that the national unemployment rate has predictive power for state unemployment rates, particularly as pertains to long-term unemployment, exhaustion of state benefits, and UI eligibility. We then use the fitted model to backcast predicted state unemployment rates before 1976.

To construct our fitted state-level series, we first estimate the relationship between the official and claims-based unemployment rates for each state i in month t with the following specification:

$$(2) \qquad UR_{i,t}^{Official} = \beta_{0,i} + \beta_{1,i}\left(UR_{i,t}^{Claims} - UR_{US,t}^{Claims}\right) + \beta_{2,i}UR_{US,t}^{Official} + \epsilon_{i,t},$$

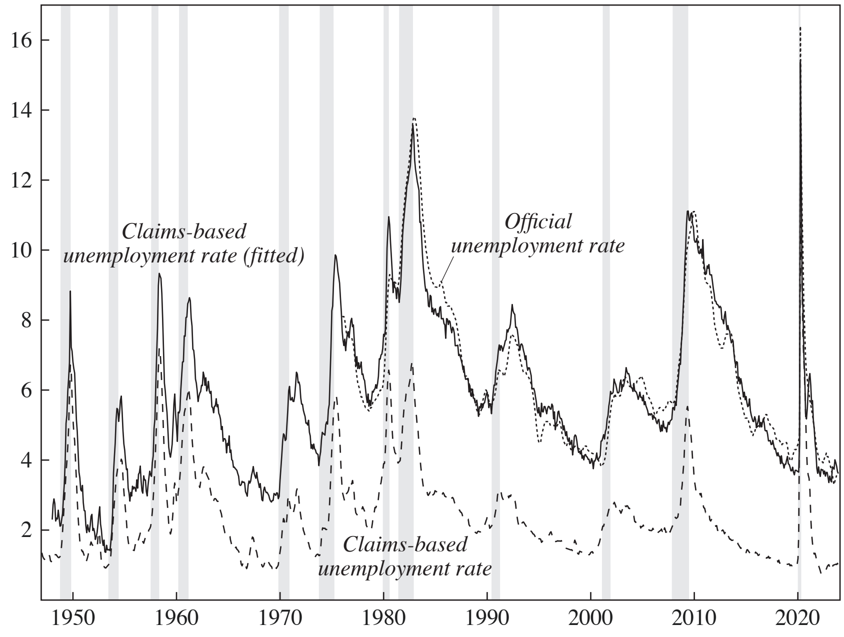
where $UR_{i,t}^{Claims} - UR_{US,t}^{Claims}$ measures the difference between the state and national claims-based unemployment rates and $UR_{US,t}^{Official}$ is the national unemployment rate. Equation (2) is estimated on data spanning January 1976 to December 2023 for each state, and we use these fitted models to generate predicted unemployment rates for January 1948 to December 1975, which are merged with model estimates for January 1976 onward.

This simple statistical model fits the state-level data extremely well, and the predicted unemployment rates capture state business cycle features that are entirely consistent with the two unemployment measures used in estimating equation (2).[19] Both correlates are highly significant predictors of a state's official unemployment rate and the average R^2 is 0.84. The average correlation coefficient of the official and predicted unemployment rates is 0.91, with a maximum of 0.98 (Indiana) and a minimum of 0.80 (Nebraska and New Mexico). Revisiting our earlier illustrative example, figure 2 plots our fitted claims-based unemployment rate for Ohio (solid black line) along with the BLS unemployment rate (dotted gray) and our unfitted claims-based unemployment rate (dashed gray) that were plotted in figure 1, panel A.[20] The fitted claims-based unemployment rate picks up on inflection points in Ohio's business cycle that are nearly identical to

19. Reassuringly, adding a covariate for the ratio of US covered unemployment to total unemployment—which, if serious, would help address the concern about UI coverage expansion discussed above—does not meaningfully improve the model fit or change the predicted series. Similarly, adding state-level controls for changes in UI policy parameters has a negligible effect on the fitted claims-based unemployment rates; see online appendix figure A.3.

20. Online appendix figure C.1 plots our fitted claims-based unemployment rates for each state along with state recession dates (gray bars) derived from them (see section III.B) and BLS unemployment rates when available.

Figure 2. Comparison of the Official and Fitted Claims-Based Unemployment Rates for Ohio



Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS.

Note: Data sample from January 1948 to December 2023. Shaded bars are NBER recession dates.

those of the official unemployment rate over 1976–2023 and to our unfitted series over the full 1948–2023 sample.

The fitted and unfitted series both have their advantages and drawbacks. One advantage of the fitted unemployment rates is that the official US unemployment rate helps anchor them, removing the level differences, but the inflection points remain virtually an identical match. The inclusion of the US unemployment rate as a regressor in equation (2) also helps to smooth the fitted claims-based unemployment data, as unemployment data are less noisy than claims data. A minor related drawback of the fitted series is that using the US unemployment rate in equation (2) modestly limits the availability of our fitted claims-based unemployment rate to January 1948, a year later than our unfitted series.

Another drawback of the fitted unemployment rates is the fact that out-of-sample observations are constructed on the assumption of a stable empirical relationship. To gauge this potential threat, we leverage state-level data available at a lower frequency to test the out-of-sample forecast of our fitted claims-based unemployment rates when feasible: in online appendix A.4, we use the CPS Annual Social and Economic Supplement (ASEC) to construct annual snapshots of state-level "unemployment rates" back to 1962 for larger states with more observations. Encouragingly, the fitted claims-based unemployment rates track the alternative ASEC-based unemployment rates quite well, both out of sample (1962–1975) and in sample (1976–1989); see online appendix figure A.6.

Reassuringly, the unfitted and fitted claims-based unemployment rate series also generate similar results when examining the timing of recessions and pace of recoveries, as discussed in section III and various robustness checks in online appendix B.2. We include both the unfitted and fitted series in our data set and let researchers determine which is more appropriate for their uses.

II. Evolving Regional Adjustments Revisited

With our claims-based unemployment rates in hand, we first use our historical data set to revisit the question of how labor markets adjust to local demand shocks, contributing new causal evidence on how those adjustments have evolved since World War II. This application builds on the seminal work of Blanchard and Katz (1992) and related work by Dao, Furceri, and Loungani (2017), among others.[21] Blanchard and Katz (1992) estimate relative employment, unemployment, and participation responses to innovations from vector autoregression (VAR) residuals over a 1978–1990 sample; they find employment responds strongly and remains persistently depressed following adverse shocks, whereas unemployment and participation see more muted, transitory responses, and thus conclude that interstate migration accounts for most of the adjustment to local demand shocks. Dao, Furceri, and Loungani (2017) revisit this question over a longer 1978–2013 sample, identifying local demand shocks using a Bartik instrument in a VAR framework; they find that labor mobility is less of an important short-run macroeconomic adjustment mechanism and state unemployment rates instead bear the brunt of the short-run adjustment. Like much of the related literature, both studies are constrained by the availability of official state unemployment, employment, and participation rates. To better

21. Several related papers study specific shocks, for example, Davis, Loungani, and Mahidhara (1997) on oil shocks and military base closures; Autor, Dorn, and Hanson (2013) on the China shock; and Yagan (2019) on the Great Recession.

understand the postwar evolution of local labor market adjustments, we estimate responses of relative employment, population, and claims-based unemployment rates to local labor demand shocks in an LP-IV framework over a much longer 1950–2019 sample and staggered subsamples.

II.A. Dispersion and Persistence of Local Adjustments

We first revisit two overarching empirical observations motivating this literature, reexamined over a longer postwar horizon. Blanchard and Katz (1992) document a wide dispersion in employment growth rates across US states; moreover, they find a high degree of persistence in states' average employment growth rates between a 1950–1970 sample and a 1970–1990 sample. In a similar spirit, Dao, Furceri, and Loungani (2017) analyze the dispersion of growth in annual employment across states over 1977–2015, finding that the standard deviation has fallen since the early 1990s; they also confirm that a high degree of persistence in employment growth and unemployment rates still holds when comparing more recent subsamples (1977–1994 and 1995–2013). The significant degree of heterogeneity and persistence in state labor market conditions would motivate interstate migration as a potentially important adjustment mechanism following local labor demand shocks.

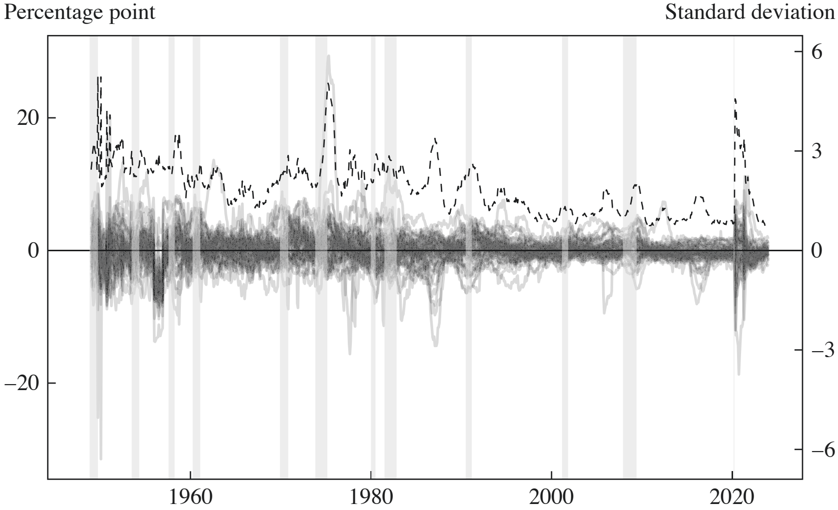
We, in turn, look at the dispersion of relative employment growth and unemployment rates across US states over a longer 1948–2023 sample. Relative employment growth (unemployment) measures each state's year-over-year growth in nonfarm payroll employment (unfitted claims-based unemployment rate) less that of the national rate. Figure 3 plots states' relative employment growth rates (panel A) and relative unemployment rates (panel B) with a solid gray line for each state (left axes) and the standard deviation across states (dashed black lines, right axes). The dispersion in relative employment growth spikes during recessions for the entire postwar sample, as Dao, Furceri, and Loungani (2017) observe since 1977, as does the dispersion in relative unemployment. But up until the COVID-19 pandemic, the degree of dispersion across states has generally trended downward throughout the postwar era, both for employment and unemployment.[22]

To quantify this trend more clearly, table 1 reports the maximum standard deviation of relative employment and unemployment from figure 3 during or within six months of each national recession, as identified by

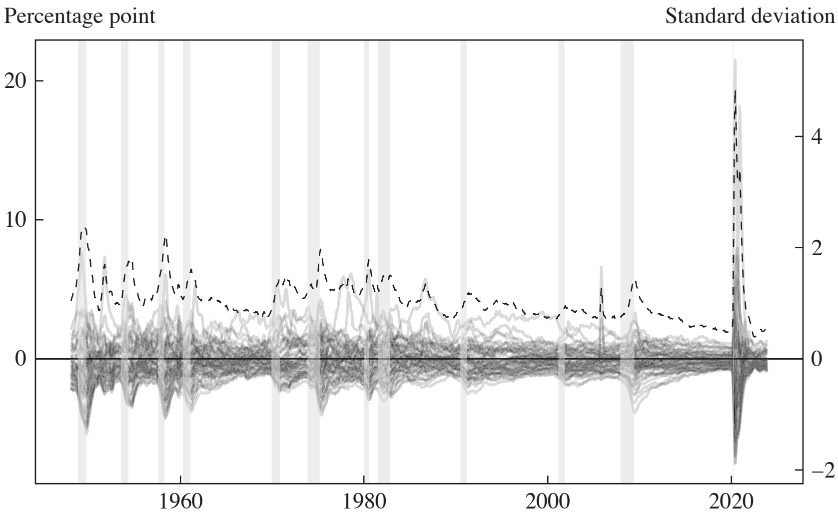
22. Stock and Watson (2010) similarly document a trend decrease in the volatility of growth in home building permits across US states over 1969–2007, particularly since the mid-1980s.

Figure 3. Dispersion of Employment Growth and Unemployment across US States

Panel A: Relative employment growth rates



Panel B: Relative claims-based unemployment rates



Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS.

Note: Data sample from January 1948 to December 2023. Relative employment growth shows each state's annual nonfarm payroll employment growth less US nonfarm payroll employment growth. Relative unemployment rates show each state's unfitted claims-based unemployment rate less the US claims-based unemployment rate. Relative employment growth and unemployment rates for each state (gray lines) are scaled on the left axis. The standard deviations of each series (dashed black line) are scaled on the right axis. Shaded gray bars are NBER recession dates.

Table 1. Maximum Labor Market Dispersion during National Recessions

Recession	*1949*	*1954*	*1958*	*1961*	*1970*	*1975*	*1980*	*1982*	*1991*	*2001*	*2009*	*2020*	*Avg.*
SD Emp.	4.7	3.1	3.4	2.6	2.8	5.5	2.9	2.9	2.6	1.4	2.0	3.9	3.1
SD CBUR	2.4	1.8	2.2	1.6	1.4	2.0	1.8	1.5	1.2	1.0	1.4	4.9	1.7

Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS.

Note: Table reports the maximum standard deviation of relative employment growth (SD Emp.) and relative claims-based unemployment rates (SD CBUR) across states during or within six months of each national recession, as plotted in figure 3. The final column reports an unweighted average across recessions excluding the pandemic (2020).

NBER. Table 1 broadly underscores a long-run trend of states increasingly experiencing more similar labor market dynamics in recessions, particularly since the early 1980s. But the results in figure 3 and table 1 document increasing similarities in state-level labor markets over the postwar era during both recessions and expansions. State-level variation notably tends to gradually decrease during relatively longer periods of economic tranquility, with the dispersion in relative unemployment trending down throughout the unusually long business cycle expansions of the 1960s and those since the start of the Great Moderation.

We also examine the persistence of labor market outcomes across states. Panel A of figure 4 compares average employment growth rates for each state in one decade relative to their growth rate in the subsequent decade, all the way from a comparison of the 1950s against the 1960s (top left) through a comparison of the 2000s against the 2010s (bottom right). A strong positive correlation is found in the early postwar decades, but since the 1970s, the relationship between average employment growth in one decade and the next has weakened considerably, highlighting less persistent differences in states' employment conditions in recent decades. Conversely, panel B of figure 4 documents a much more persistently positive relationship between states' relative claims-based unemployment rates across decades.[23] Both the employment and unemployment persistence figures also show less dispersion in labor market conditions in more recent decades, consistent with the results in figure 3. Broadly speaking, our evidence of a diminishing persistence in employment growth and less dispersed labor market outcomes suggests less scope for migration as an adjustment to local labor market shocks in more recent decades.

II.B. Adjustments to Bartik Industry Share Shocks

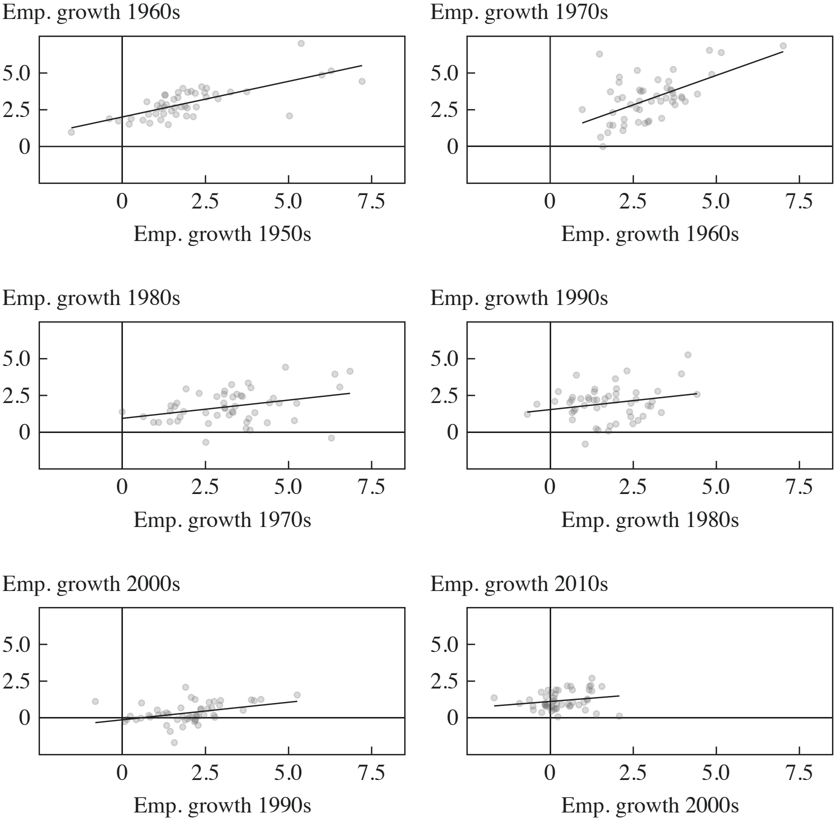
In this section, we more formally analyze the local labor market responses to labor demand shocks and assess how these responses have changed over the postwar era. For our local labor market demand shock, we take a similar Bartik instrument approach to identification as used by Dao, Furceri, and Loungani (2017).[24] To capture each state's exposure to national labor

23. Neumann and Topel (1991) document a similar persistence in annual relative state IURs over intervals spanning 1950–1985, consistent with our longer analysis of monthly claims-based unemployment rate data and contrary to the evidence presented by Blanchard and Katz (1992).

24. Davis, Loungani, and Mahidhara (1997) take a similar approach to constructing state industry mixes of employment interacted with national employment growth rates as one of their studied local demand shocks.

Figure 4. Changes in Log Employment and Unemployment by Decade across US States

Panel A: Employment growth rates



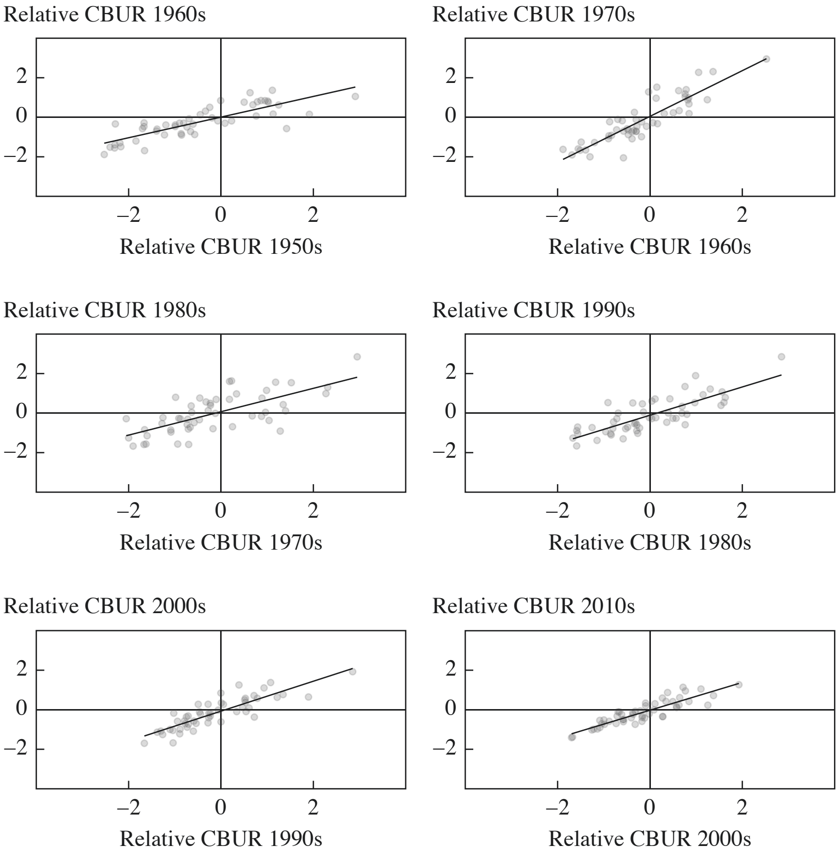
demand shocks, we construct a relative industry mix variable for each state's annual personal income growth weighted by industrial composition, measured relative to the national average for industry-weighted personal income growth. For state i, the industry mix variable in year t, $imix_{i,t}$, is constructed as a weighted share of personal income growth across J industries:

$$imix_{i,t} = \sum_{j=1}^{J} \left[\bar{\theta}_{i,j,t} \Delta \ln\left(\bar{\iota}_{-i,j,t} \right) \right]. \tag{3}$$

The weights $\bar{\theta}_{i,j,t}$ in equation (3) reflect state i's share of personal income growth in industry j in year t, taken as a five-year moving average, which are used to scale $\Delta\ln(\bar{\iota}_{-i,j,t})$, the annual growth of log personal income in

Figure 4. Changes in Log Employment and Unemployment by Decade across US States (*Continued*)

Panel B: Relative claims-based unemployment rates



Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS.

Note: Data sample from 1950 to 2010. Panel A plots each state's average annualized log employment growth in one decade against that of the next decade. Panel B plots each state's average relative unfitted claims-based unemployment rate in one decade against that of the next decade.

industry j for all states excluding i. Our relative industry income mix variable, $rimix_{i,t}$, subtracts the national average from each state's industrial mix of personal income growth.

Dao, Furceri, and Loungani (2017) construct a version of this state-level industry mix Bartik instrument from equation (3) using nonfarm private sector employment by industry from the Bureau of Economic Analysis (BEA)

Regional Economic Accounts (REA), based on twenty industries at the two-digit Standard Industrial Classification (SIC)/North American Industry Classification System (NAICS) industry level. The REA data on employment by industry, however, are only available starting in 1969. We instead use REA data on annual personal income by major component and earnings by industry because they are available for most states back to 1929, allowing us to construct a Bartik instrument for the full sample of our claims-based unemployment rates.[25] Our relative industry income mix variable is based on nine industry groups that can be consistently constructed across the NAICS, SIC, and BEA's historical industry classifications. While less refined in terms of industry exposure, our use of earnings as opposed to employment is advantageous for capturing labor demand shocks through both the extensive and intensive margins; changes in hours worked might be more relevant than changes in employment following certain national shocks.

We construct our Bartik instrument from earnings by industry using BEA's historical industry classifications up through 1957, SIC for 1958–2001, and NAICS starting in 2002. Because the transition from SIC to NAICS industry codes resulted in a large reclassification of workers from manufacturing to services employment (Fort and Klimek 2018), in our baseline analysis below we drop all observations of our Bartik instrument over 2000–2003 to avoid any potential industry contamination around the SIC/NAICS transition in 2002 from the five-year centered moving average; this approach is equivalent to constructing separate instruments predating the NAICS era (up to 1999) and for the NAICS era (2004 onward).[26]

Our choice of labor market outcome variables is also motivated by historical data availability. Blanchard and Katz (1992) and Dao, Furceri, and Loungani (2017) estimate VAR systems using changes in relative (log) employment growth, the relative (log) employment rate, and the relative (log) participation rate, and then back out the implied responses

25. Data for Alaska and Hawaii start in 1950, and data are available back to 1929 for all other states.

26. We thank Christopher House, Andrea Foschi, Linda Tesar, and Christian Pröbsting for kindly flagging this concern about potential measurement error stemming from the SIC/NAICS transition in such Bartik instruments. The construction by Dao, Furceri, and Loungani (2017) of a similar Bartik instrument mixes employment across the two industry classifications around this time, which may be biasing their related results; see discussion in online appendix B.2.

of unemployment and population growth.[27] But like the official state unemployment rates, monthly state-level participation and employment rates are only available back to 1976. We analyze dynamics of our claims-based unemployment rate, nonfarm payroll employment, and total civilian population because they are all available at the state level back to the 1940s. Dao, Furceri, and Loungani (2017) argue that, in response to local demand shocks, changes in the civilian working-age population—which they use as a proxy for migration—should primarily be driven by net migration, as adult mortality, incarceration, and immigration from abroad are unlikely to respond quickly to local demand shocks; in the same vein, we expect that the response of total population should also largely reflect net migration, as births and child mortality are also unlikely to respond strongly or quickly to local labor demand shocks.

We estimate state labor market adjustments to local demand shocks in the following reduced form LP-IV panel regression framework:

$$\Delta Y_{i,t+h} = \alpha_i + \gamma_t + \beta_h rimix_{i,t} + \phi_h\left(L\right)\boldsymbol{Z}_{i,t-1} + \epsilon_{i,t+h}, \tag{4}$$

where α_i and γ_t are state and year fixed effects, respectively, and $rimix_{i,t}$ is the relative industry income mix Bartik instrument for state i in year t. For dependent variable of interest $\Delta Y_{i,t+h}$, we rotate in the cumulative change in relative unemployment ($\Delta Y_{i,t+h} = \tilde{Y}_{i,t+h} - \tilde{Y}_{i,,t-1}$), the cumulative log point change in relative employment [$\Delta Y_{i,t+h} = \ln(\tilde{Y}_{i,t+h}) - \ln(\tilde{Y}_{i,t-1})$] and the cumulative log point change in relative population. The respective relative variables are constructed by subtracting the claims-based unemployment rate for the United States from the claims-based unemployment rate for state i, subtracting log employment for the United States from log employment for state i, and subtracting log population for the United States from log population for state i. Regardless of which variable is being rotated in for $\Delta Y_{i,t+h}$, $\boldsymbol{Z}_{i,t-1}$ is a vector of lagged controls containing first differences of the

27. While we do not study employment rates due to historical data limitations, several other papers have found persistent local employment rate responses more in keeping with our results below, contrary to the transitory response of relative employment rates or participation rates documented by Blanchard and Katz (1992) and Dao, Furceri, and Loungani (2017). For instance, Autor, Dorn, and Hanson (2013) find persistent local labor market responses in areas more exposed to increased Chinese import demand over 1990–2007. And studying state-level labor market hysteresis following the Great Recession, Yagan (2019) finds a high degree of persistence in both employment and employment rates for states more exposed to the recession.

relative claims-based unemployment rate and log first differences of relative employment and population for each state, to mop up any state-specific labor market trends not absorbed by the state and time fixed effects. In keeping with the two-lag annual VAR specifications of Blanchard and Katz (1992) and Dao, Furceri, and Loungani (2017), $\phi_h(L)$ is a lag polynomial of order two.

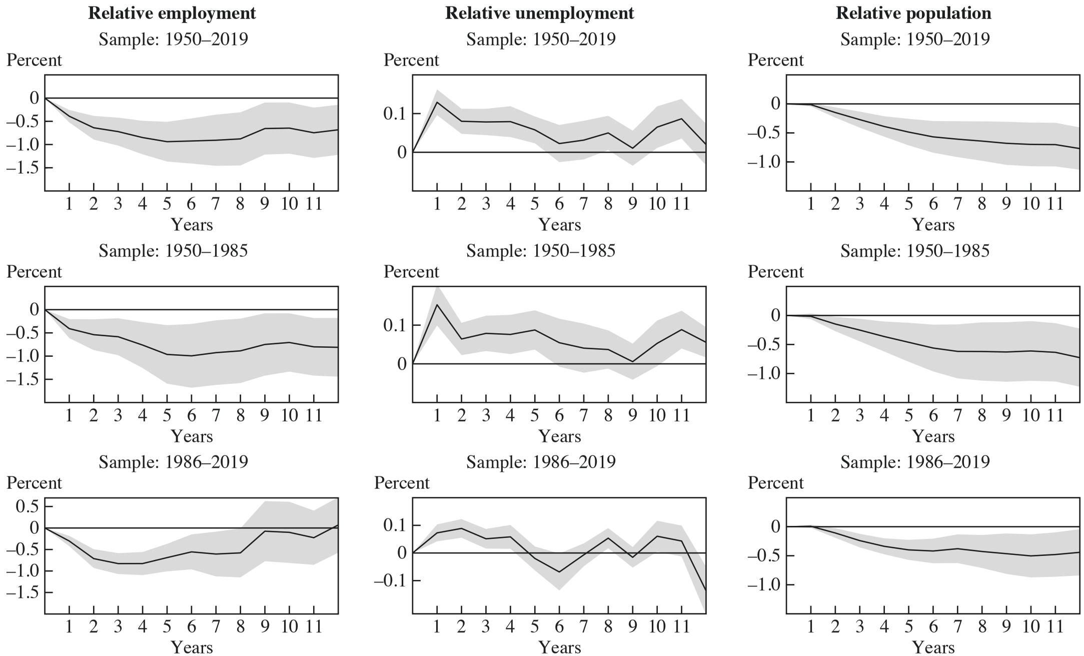
For our benchmark analysis, we first estimate the local projections in equation (4) over an annual sample of 1950–2019, with separate regressions for each forecast horizon $h \in \{0, 1, \ldots, 11\}$.[28] The estimated sequence $\left\{\widehat{\beta_h}\right\}_{h=0}^{11}$ traces out the dynamic impulse response function for the cumulative changes in relative labor market outcomes over the twelve-year forecast horizon in response to a −1 percentage point shock to state i's personal income growth, given its industrial composition, relative to the national average growth rate. The cumulative log point responses of relative employment and population reflect growth in state i less average national growth over the same horizon, so impulse responses for all dependent variables are measured in percentage points.

Because local labor market adjustments can be quite sensitive to sample selection, we also estimate the same regressions over two evenly split subsamples: 1950–1985, reflecting the early postwar era through the Great Inflation, and 1986–2019, capturing the Great Moderation through recovery from the Great Recession (deliberately excluding the pandemic). The top row of figure 5 depicts the impulse responses of relative (log) employment (left), relative unemployment (middle), and relative (log) population (right) in response to the Bartik demand shocks over the full sample, along with shaded 95 percent confidence bands constructed from robust standard errors, clustered at the state level. The middle and bottom rows of figure 5 depict the responses of the same relative variables estimated over the earlier and more recent subsamples, respectively.

Over the full postwar sample, the estimated response of relative employment growth to an adverse local labor market demand shock shows a gradual but persistent decline, peaking 0.94 percentage points below national growth after five years; the decrease is significant at the 95 percent confidence level throughout the forecast horizon. The response of relative unemployment is much more immediate, with a peak increase of 0.13 percentage

28. Our unfitted claims-based unemployment rate data begin in 1947, but we lose two burn-in years in our sample due to the first-differenced lagged controls and another to the differenced dependent variables. The choice of a twelve-year impulse response horizon is in keeping with Blanchard and Katz (1992) and Dao, Furceri, and Loungani (2017).

Figure 5. Local Labor Market Responses to Bartik Demand Shocks



Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS.

Note: Data sample from 1950 to 2019 and subsamples. Figures depict the impulse responses of relative labor market variables estimated by the local projections in equation (4), with the $rimix_{i,t}$ instrument scaled to a 1 percentage point decrease in state i's personal income growth relative to national average growth and dropping observations for 2000–2003. Shaded bands denote 95 percent confidence intervals.

points after one year; the jump in unemployment is less persistent than the decline in employment, and the null hypothesis of no effect on unemployment cannot be rejected at conventional levels of significance for much of the second half of the impulse response horizon. Lastly, the negative response of relative population is even more gradual and persistent than that of employment, declining as much as 0.77 percentage points by the end of the twelve-year forecast horizon; save the first year, when there is no response on impact, the decline in population is consistently significant at the 95 percent confidence level.

Our LP-IV impulse responses for the full sample qualitatively resemble the main Bartik IV estimates of Dao, Furceri, and Loungani (2017), despite varying regression frameworks and samples.[29] Across both, unemployment sees the greatest response on impact, with a peak effect after one year; employment sees a more gradual but fairly persistent decline; and population sees a more gradual and more persistent decline than employment. Moreover, our benchmark impulse response estimates are far more similar to the Dao, Furceri, and Loungani (2017) IV estimates than their OLS estimates, intended as more in keeping with the earlier Blanchard and Katz (1992) estimates.

Looking to the bottom two rows of figure 5, it is clear that state labor market adjustments to local shocks have weakened in more recent decades. The response of relative employment growth to an adverse local labor demand shock is somewhat stronger in the earlier postwar sample, falling 0.99 percentage points after six years, whereas the drop in employment is smaller and less gradual in the more recent sample, with peak decline of 0.83 percentage points after three years; and the decline in employment remains highly persistent and significant at the 95 percent level throughout the forecast horizon when estimated over 1950–1985, whereas the decline is transitory when estimated over 1986–2019, with point estimates roughly reverting to zero after a decade.

Similarly, an immediate jump in unemployment is even more pronounced when estimated over 1950–1985, rising by 0.15 percentage points after one year, than when estimated over the entire sample but again generally follows a qualitatively similar path as the estimates for 1950–2019. Conversely, the rise in unemployment estimated over 1986–2019 is more muted, with a peak increase of 0.09 percentage points after two years, and less persistent.

29. Dao, Furceri, and Loungani (2017) do not plot confidence intervals in their figures, hence our inability to speak to comparisons of the statistical significance of our two sets of impulse responses.

But the most notable difference between labor market adjustment margins between the two subsamples are those of population. Estimated over 1950–1985, relative population sees a gradual and highly persistent decline quite similar to the impulse response estimated over the full sample, falling 0.73 percentage points by the end of the forecast horizon. Estimating equation (4) over the 1986–2019 sample, we instead see a peak relative population response of 0.50 percentage point decline after ten years, moderating to 0.44 percentage points by the end of the forecast horizon. Dao, Furceri, and Loungani (2017) find an even sharper attenuation of relative population responses when comparing post-1990 subsamples with a 1978–1990 subsample, though their Bartik instrument may reflect mismeasurement from the SIC/NAICS transition in more recent samples.[30] Regardless, our results place such a trend in much longer historical context, and both sets of results may mirror a broader trend of decreasing internal migration rates in recent decades, particularly since the early 1980s (Molloy, Smith, and Wozniak 2011).

In the more recent subsample, the still sizable but less persistent response of relative employment, weaker response of relative unemployment, and more muted response of relative population all seem consistent with labor force participation playing more of a role in adjustment and employment recovering because of decreased out-migration—consistent with recent evidence that nonparticipation is a key adjustment margin (Elsby, Hobijn, and Şahin 2015; Yagan 2019).

Given the sensitivity of local labor market adjustments to choices about sample, in online appendix B.2 we estimate the local projections in equation (4) as rolling regressions over staggered twenty-five-year estimation samples. Online appendix figure B.6 shows that the headline results in figure 5 for our postwar subsample split are broadly robust to more refined, staggered subsamples: across relative employment, unemployment, and population we consistently see a substantial attenuation of impulse responses estimated starting in the mid-1980s or later, with these peak effects shrinking roughly 45–65 percent from those estimated over earlier decades.

In the online appendix, we also disaggregate our $rimix_{i,t}$ Bartik instrument into "very positive" and "very negative" shocks, defined as above-average positive and below-average negative shocks, respectively, and document

30. Online appendix figure B.5 replicates figure 5 without dropping our Bartik instrument for 2000–2003, which shows even greater attenuation of the relative population response between the 1950–1985 and 1986–2019 subsamples, more similar to the relative population responses reported by Dao, Furceri, and Loungani (2017).

nonlinear effect of local shocks, not just sign but in magnitude; see online appendix figure B.7. The key takeaway from our disaggregation is that there are fewer "very positive" and "very negative" shocks in the 1986–2019 sample than the 1950–1985 sample, and the mean (absolute) values are smaller in the later sample than the earlier sample.[31] Fewer and relatively smaller "large" relative labor market shocks in more recent decades correspond with the declining dispersion in states' relative employment growth and unemployment rates seen in figure 3, all wholly consistent with a smaller migration response in recent decades: the improvements in labor market conditions that could be achieved by migration appear to have diminished in more recent decades as states look increasingly similar across the national business cycle.

III. Unemployment Recoveries from Recessions

Studies of state or regional business cycles often difference out the national business cycle—as our analysis above did—for good reasons, for example, stationarity of outcome variables, identifying variation from Bartik instruments, and so on. But the US business cycle is surely influencing the evolution of states' labor market adjustments analyzed in section II. Moreover, evidence from states can help shed light on the evolution of the aggregate US business cycle. In this section, we use our historical measures of unemployment to examine various features of postwar US business cycles and the evolution of unemployment dynamics at both the state and national level.

We first use our claims-based unemployment rates to identify business cycle peaks and troughs for the United States and for all fifty states; we show that these recession dates line up reasonably well with inflection points estimated from the official US unemployment rate and existing estimates of state recession probabilities in overlapping samples, respectively. We then use these recession dates and our new data set to study the rate and degree to which unemployment recovers following recessions. Most related research on unemployment dynamics focuses on the national business cycle, but our historical claims-based unemployment rates allow us to study heterogeneity across states, exploiting a much larger postwar sample of observations. In particular, we explore the evolving dispersion of states'

31. For $rimix_{i,t}^{++}$, there are 346 observations (mean of 1.3 percentage points) in the early sample and 207 observations (mean of 1.0 percentage points) in the later sample. For $rimix_{i,t}^{--}$, there are 314 observations (mean of −1.4 percentage points) in the early sample and 163 observations (mean of −1.1 percentage points) in the later sample.

unemployment recovery rates as well as changes in unemployment relative to state-specific business cycle troughs.

III.A. Recession Dating for the US Business Cycle

To analyze the speed and dispersion of unemployment recoveries, we must first choose a chronology of business cycle inflection points.[32] There are various approaches to identifying peaks and troughs in the business cycle; see Romer and Romer (2019) for an overview. We adopt the relatively simple, unemployment-based recession dating algorithm proposed in Dupraz, Nakamura, and Steinsson (2024), which generates a close match to the NBER recession dates.[33] The Dupraz, Nakamura, and Steinsson (2024) recession dating algorithm identifies local minima and maxima of the unemployment rate, ignoring low-frequency variation, similar in spirit to the Bry and Boschan (1971) algorithm or the unemployment-based Sahm (2019) rule; see online appendix A.5 for an overview of the Dupraz, Nakamura, and Steinsson (2024) algorithm.[34] Table 2 reports national business cycle peak and trough dates identified by the algorithm from our US claims-based unemployment rate as well as those from the BLS unemployment rate, along with NBER recession dates as a benchmark.

The two unemployment-based chronologies of recession dates generate a relatively consistent match with one another. The peaks and troughs identified from the claims-based unemployment rate generally occur earlier than those generated from the official unemployment rate, with an average absolute discrepancy of 3.8 months for troughs versus 5.8 months for peaks. The UI claims we use in constructing our claims-based unemployment rates are faster to pick up changes in the labor market—in particular, IC are a leading economic indicator—than the official unemployment rate,

32. The common alternative to using chronologies is estimating a Markov regime-switching model, first popularized by Hamilton (1989), in which turning points are unobserved latent variables; the model produces posterior probabilities that a given period is an inflection point, and hence recession probabilities. A chronology of inflection points is far more tractable for estimating recovery speeds and for comparisons with the recent literature on national recoveries.

33. The Dupraz, Nakamura, and Steinsson (2024) algorithm also identifies peaks and troughs in US business cycles that are nearly identical to the chronology in Hall and Kudlyak (2020, 2022) based on observed peaks and troughs in the US unemployment rate.

34. As a robustness check, we also estimate state recession peaks and troughs using the Bry and Boschan (1971) algorithm, which generates similar results. The Sahm (2019) rule heuristic for identifying recessions is based on the three-month moving average of the US unemployment rate rising at least 0.5 percentage points above its preceding twelve-month low, which could also be adapted to state unemployment rates.

Table 2. Business Cycle Peaks and Troughs

	NBER		*Dupraz, Nakamura, and Steinsson (2024) dating algorithm: CBUR*		*Dupraz, Nakamura, and Steinsson (2024) dating algorithm: Official UR*	
	Peak	*Trough*	*Peak*	*Trough*	*Peak*	*Trough*
1	Nov. 1948	Oct. 1949	[Dec. 1948]	Oct. 1949	[Jan. 1948]	Oct. 1949
2	[July 1953]	May 1954	Apr. 1953	Sep. 1954	May 1953	Sep. 1954
3	Aug. 1957	Apr. 1958	Dec. 1955	May 1958	Mar. 1957	July 1958
4	Apr. 1960	Feb. 1961	June 1959	Mar. 1961	Feb. 1960	May 1961
5	Dec. 1969	Nov. 1970	June 1969	Nov. 1970	Sep. 1968	Dec. 1970
6	Nov. 1973	Mar. 1975	Apr. 1973	May 1975	Oct. 1973	May 1975
7a	Jan. 1980	July 1980	Nov. 1978	July 1980	May 1979	
7b	July 1981	Nov. 1982	June 1981	Oct. 1982		Nov. 1982
8	July 1990	Mar. 1991	Nov. 1988	Mar. 1991	Mar. 1989	June 1992
9	Mar. 2001	Nov. 2001	Apr. 2000	Mar. 2002	Apr. 2000	June 2003
10	Dec. 2007	June 2009	Apr. 2006	May 2009	Oct. 2006	Oct. 2009
11	[Feb. 2020]	Apr. 2020	June 2019	May 2020	Sep. 2019	Apr. 2020

Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS.

Note: Recession dates for the claims-based unemployment rate (CBUR) and official unemployment rate (UR) are generated using the Dupraz, Nakamura, and Steinsson (2024) algorithm, setting parameter $X = 1.5$ for the official UR and $X = 1.0$ for the CBUR; see online appendix A.5 for details. For the NBER recession dates, the peaks in July 1953 and February 2020 are bracketed to note that the identified *quarterly* peak occurred earlier, in 1953:Q2 and 2019:Q4, respectively. For the algorithm, the peaks in December 1948 and January 1948 are bracketed to note that the algorithm cannot identify those peaks due to data limitations; both dates are hard-coded based on minima during available samples.

a lagging economic indicator. Online appendix figure B.8 plots cross-correlograms for the official US unemployment rate versus our claims-based unemployment rate as well as the IUR; these figures highlight that, in addition to being highly correlated with the official unemployment rate, both claims-based indicators tend to slightly lead the official unemployment rate—consistent with business cycle peaks and troughs being identified slightly earlier when using our claims-based unemployment rate instead of the official unemployment rate, as seen in table 2.

Unsurprisingly, the unemployment-based recession dates align better for troughs than peaks. A challenge with the Dupraz, Nakamura, and Steinsson (2024) algorithm is a sensitivity to "flat peaks" in economic activity, that is, trying to identify local minima around low and relatively stable unemployment rates late into business cycle expansions, which is never an issue for troughs. The worst peak match is the end of the recovery from the 1953–1954 recession, where the peak dates from the unemployment-based chronologies are fifteen months apart—a dating discrepancy easily understood by looking at figure 1, panel B, which shows an unusually flat peak in US unemployment. The official unemployment rate (gray dashed line) reaches close to its local minimum in 1955 but jumps around before reaching a minimum in March 1957, whereas the claims-based unemployment rate (black line) hits its local minimum in December 1955 and is slightly trending upward into 1957. If we discard this extreme case, the average discrepancy between the peak dates is 4.8 months, roughly in line with the average absolute discrepancy in troughs, implying comparable recession recovery durations on average.

Overall, the unemployment-based recession dates also generate a relatively consistent match with the NBER business cycle dates. One notable difference between the two unemployment-based recession dates is that the claims-based unemployment rate series identifies a double-dip recession in the early 1980s, spot on with the July 1980–July 1981 recovery identified by NBER, but only a single, longer recession is identified from the official unemployment rate.[35] Looking back to figure 1, this divergence is again easily understood: there is only a modest decline in the official unemployment series in late 1980 and early 1981 but a much more pronounced dip in our claims-based unemployment series. Our US claims-based unemployment rate also generates a closer match to the NBER trough dates than does the official unemployment rate: ignoring the initial 1980 recession, the average absolute difference between the claims-based and NBER trough dates is 1.4 months, versus 4.6 months between the official unemployment rate and NBER trough dates. Both the claims-based and official unemployment rate series do a worse job matching the NBER peaks than matching the NBER troughs, again reflecting the flat-peak challenge with the Dupraz, Nakamura, and Steinsson (2024) algorithm; the claims-based and official unemployment rate recession dates have an average absolute discrepancy of 11.9 months and 7.9 months, respectively, from the NBER peaks.

In our baseline analysis below, we employ the recession dates inferred from the official US unemployment rate as a better crosswalk with the existing literature. Online appendix B.3 provides additional results using recession dates inferred from our claims-based unemployment rates; our headline results are robust to choices about chronologies of national recession dates.

III.B. Recession Dating for US States

To understand the evolution of state-level business cycles, we next need business cycle peaks and troughs at the state level, and we construct them

35. The chronology of recession dates identified by Hall and Kudlyak (2020, 2022) from the US unemployment rate similarly does not identify a double-dip recession in the early 1980s.

by applying the Dupraz, Nakamura, and Steinsson (2024) recession-dating algorithm to our fitted claims-based unemployment rate for each state.[36] Online appendix figure C.1 depicts our claims-based unemployment rates (blue lines) and the state recession dates derived from them (gray bars) for every state. As a validation exercise, we compare our claims-based peak and trough dates for US states with the state recession probabilities estimated by Owyang, Piger, and Wall (2005) using a Markov regime-switching model; they produce estimates of state recession probabilities for February 1979–June 2002, a sample limited by the availability of the state coincident indexes, which in turn are limited by the unavailability of the BLS state unemployment rates before 1976.[37] Online appendix figure C.1 also depicts the Owyang, Piger, and Wall (2005) state recession probabilities for this subsample (red lines).

Broadly speaking, the crosswalk suggests that our claims-based unemployment rates identify similar business cycle dynamics for most states, particularly larger ones, during the overlapping 1979–2002 sample; the similarities and differences between our state-level recession dates and the recession probabilities in Owyang, Piger, and Wall (2005) are discussed in more detail in online appendix C. Some differences are to be expected. Markov-switching models and the Dupraz, Nakamura, and Steinsson (2024) algorithm identify fundamentally different objects, and state-level coincident indexes are a related but broader measure of economic activity than state unemployment rates, our exclusive focus in identifying recession dates. Neither approach is right or wrong per se. But online appendix figure C.1 underscores a drawback of using the Markov regime-switching approach for studying recovery rates: our recession dates exhibit fewer

36. It is less obvious how to appropriately set the parameter $\boldsymbol{X}$ for the Dupraz, Nakamura, and Steinsson (2024) algorithm for states, which have more varied amplitudes of unemployment than the nation. For the national (unfitted) claims-based unemployment rate we set $\boldsymbol{X} = 1.0$, which generates a good match with NBER recession dates (reported in table 2). For states, we compute the ratio of the state-level and national claims-based unemployment rates and scale each state's $\boldsymbol{X}$ parameter accordingly, and then rescale these down by 25 percent to be conservative; see online appendix A.5 for details.

37. The Federal Reserve Bank of Philadelphia produces up-to-date monthly state coincident indexes using the model of Crone and Clayton-Matthews (2005), but data are similarly only available starting in January 1979 or later. The coincident indexes are estimated from four state-level variables: nonfarm payroll employment, average hours worked of production workers in manufacturing, the official state unemployment rate, and real wage and salary disbursements; see Federal Reserve Bank of Philadelphia, "State Coincident Indexes," https://www.philadelphiafed.org/surveys-and-data/regional-economic-analysis/state-coincident-indexes.

erratic, short-lived recessionary spikes or dubiously long recessionary periods, and no judgment is required regarding a cutoff for recession probabilities to identify recovery dates and durations. The principal advantage to our approach, however, is the ability to identify inflection points in state business cycles for more than thirty additional years when using our claims-based unemployment rate series instead of existing off-the-shelf state coincident indexes. It would be possible to construct backdated coincident indexes using our new data set and estimate state recession probabilities over a longer horizon, but we leave that for future research.

III.C. National and State Unemployment Recovery Rates

With these recession dates in hand, we use our new claims-based unemployment rates to examine the evolving pace of economic recoveries at both the state and national level as well as the dispersion of recovery rates across states. Following the general approach in Hall and Kudlyak (2020), we compute the pace of recovery in unemployment as the mean decline in the log unemployment rate, UR_t, over the recovery period, defined as:

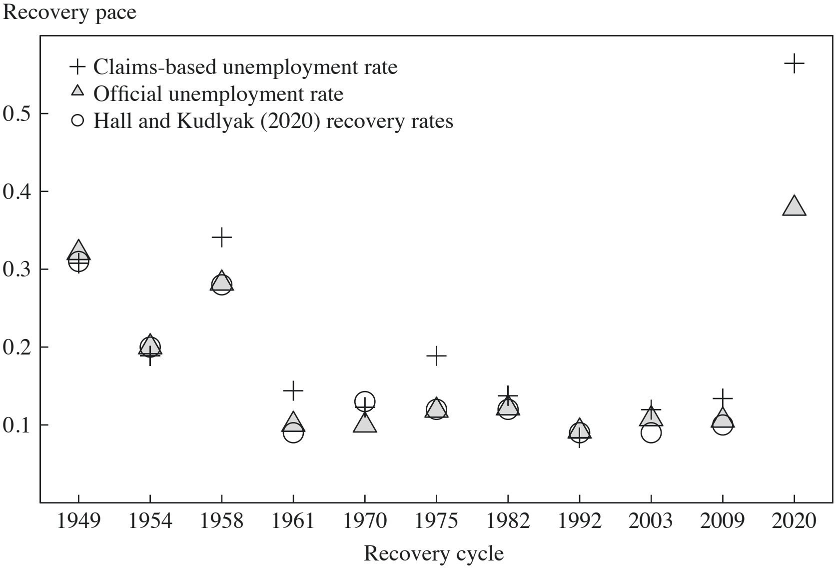
$$\text{(5)} \qquad \text{Recovery Pace} = -12 \cdot \left(\log UR_0 - \log UR_T\right) \Big/ T\,.$$

Equation (5) calculates the average annualized percentage decline in the unemployment rate from its maxima at the end of a recession (recovery starting at month zero) to its minima at the end of the ensuing expansion (recovery ending at month T).

Figure 6 depicts the national recovery rates for the official US unemployment rate (triangles) and our claims-based unemployment rate (crosses) for eleven postwar expansions, calculated from the recession dates in table 2 derived from the US unemployment rate. As a benchmark, figure 6 also replicates the national unemployment recovery rates from figure 3 of Hall and Kudlyak (2020) for the first ten recoveries (circles), constructed from their unemployment-based chronology of recession dates.[38] Figure 6 shows that national unemployment recovery rates have decelerated markedly since the 1950s and roughly stabilized starting in the 1960s, at least up until

38. In a subsequent version of their paper, Hall and Kudlyak (2022) revise their methodology for estimating recovery rates, such that equation (5) is a nested case when log unemployment is a random walk—a reasonable approximation of reality, as the autocorrelation of the log US unemployment rate and US claims-based unemployment rate both exceed 0.98. We view equation (5) as the appropriate descriptive statistic for average recovery rates of unemployment, and a far more tractable approach for calculating 500-plus recovery rates than their bootstrapped estimation of ten recoveries.

Figure 6. National Recovery Rates of Official and Claims-Based Unemployment Rates



Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS and data reproduced from Hall and Kudlyak (2020), which were kindly provided by Kudlyak.

Note: Recovery dates are estimated from the official US unemployment rate using the Dupraz, Nakamura, and Steinsson (2024) algorithm; see table 2 for dates. The recovery from the pandemic recession is dated from the trough in April 2020 to a peak in December 2023 (the end of our sample). Recovery from the 1980 recession is excluded because that expansion is only identified from the claims-based unemployment rate series and recovery is cut short by the more severe 1981–1982 double-dip recession.

the pandemic. Encouragingly, our claims-based unemployment rate generates very similar recovery rates as the official unemployment rate—using either our recession dates or the Hall and Kudlyak (2020) chronology—for the first ten postwar recessions; our series also identifies the same structural break in recovery rates between the 1957–1958 and 1960–1961 recessions. It is important to emphasize that our US claims-based unemployment rate is not fitted using the official unemployment rate as in equation (2)—it is computed from unfitted claims data; see equation (1).

The choice of recession dates influences the calculation of recovery speeds, both in terms of the log point change in the unemployment rate and potentially in the duration of the recession, as underscored by several slight differences in the US unemployment recovery rates when using the recession dates based on Dupraz, Nakamura, and Steinsson (2024) versus dates from Hall and Kudlyak (2020, 2022). As a robustness check, online appendix figure B.9 replicates figure 6 using recession dates estimated from our US claims-based unemployment rate (also in table 2) instead of those derived from the BLS unemployment rate. The broad trends of a marked deceleration in unemployment recoveries since the 1950s and more stable, uniform recovery rates over the last sixty years hold using either set of recession dates.[39]

The only major divergence between the two recovery rates comes after the pandemic, when the claims-based unemployment rate shows a much faster recovery rate than the official unemployment rate, as would be expected from figure 1. Both series see a comparable spike in March–May 2020, but during the ensuing recovery the claims-based unemployment rate quickly falls below pre-pandemic levels to record lows, whereas the official unemployment rate did not recover to its pre-pandemic rate until July 2022. The differential degrees of recovery are amplified into fast, more divergent recovery rates by the historically short time to recovery.

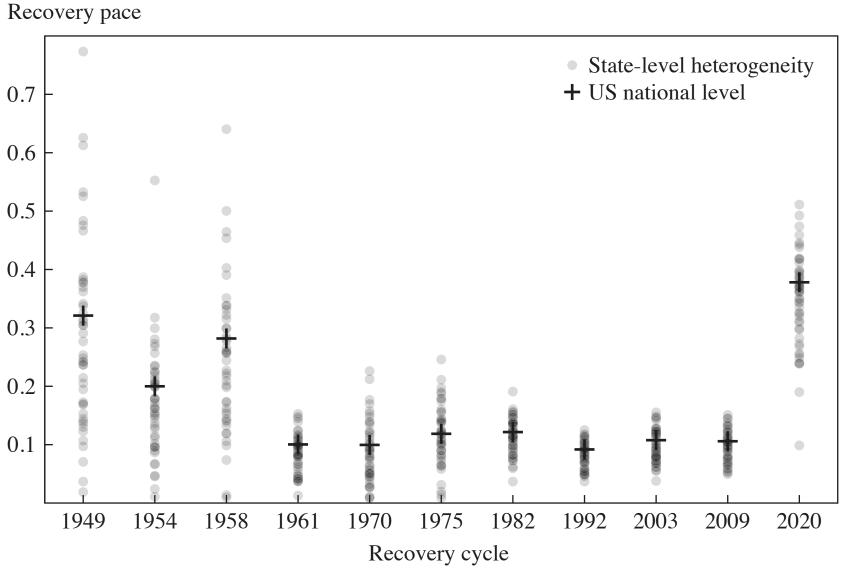
We next explore the pace of economic recoveries across states for the same eleven recessions, using our fitted claims-based unemployment rates. Figure 7 plots the state-level recovery rates as circles along with crosses depicting recovery rates for the US unemployment rate (previously plotted in figure 6).[40] One striking feature of these data is that faster recoveries tend to be associated with much more dispersion in the pace of states' recoveries: this was true during the faster recoveries from the early postwar recessions of 1948–1949, 1953–1954, and 1957–1958, and this dynamic reemerged in the rapid pandemic recovery, albeit likely for different reasons discussed below.

To display this association more clearly, online appendix figure B.10 plots the national recovery pace against the standard deviation of state-level recovery rates, which displays a clear increasing relationship. Again, faster national recoveries tend to be ones where states experience very different outcomes, and states experience rather similar unemployment recoveries during slower national recoveries throughout the 1960s–2010s. The deceleration and convergence in states' unemployment recovery rates depicted in figure 7, most prominent since the 1982 recession, corresponds with the evidence from figure 5 showing a shift from more immediate jumps in

39. If anything, the deceleration in recovery rates since the 1950s is even more pronounced when using recovery dates estimated from our claims-based unemployment rates.

40. The official US unemployment rate is the better benchmark for recovery rates estimated from our fitted claims-based unemployment rates than our US claims-based recovery rate; see equation (2).

Figure 7. State-Level Recovery Rates of Claims-Based Unemployment Rates



Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS.

Note: Recovery dates are estimated from the official US unemployment rate using the Dupraz, Nakamura, and Steinsson (2024) algorithm; see table 2. Recovery from the 1980 recession is excluded and recovery from the pandemic is hard-coded to a peak in December 2023; see note for figure 6. Recovery rates are negative for a few states, that is, their unemployment rate rose during the national recovery, but only nonnegative recovery rates are plotted.

relative unemployment and faster ensuing recoveries early in the postwar era to more gradual, persistent increases in relative unemployment since the start of the Great Moderation.

Echoing the evidence in figure 3, these results also underscore that labor market conditions are increasingly similar across states in more recent recessions, possibly reducing the job prospects workers can achieve via migration. The rapidness of the early postwar recoveries may be driven, at least in part, by a migration response that subsequently weakened: larger differences across local labor markets may induce greater labor mobility, quickening the national adjustment to recessionary shocks. Saks and Wozniak (2011) document that US migration rates are generally procyclical but note that the recessions of 1957–1958 and 1960–1961 are the only postwar exceptions in which interstate migration rates instead rose during a recession; the 1948–1949 recession was also soon followed by the largest one-year jump in interstate migration rates in their postwar sample. Prior to the pandemic, the 1949 and 1958 recessions also saw the fastest national unemployment recovery rates and the greatest dispersion of recovery rates across states, as seen in figure 7.

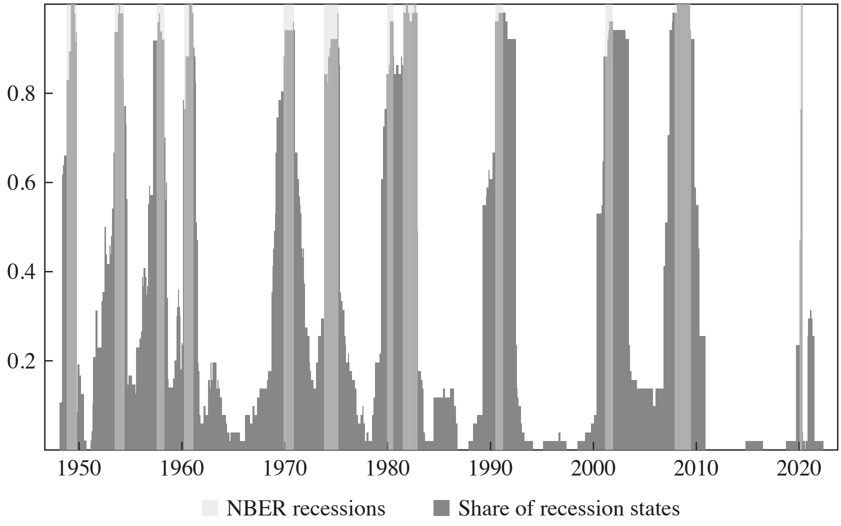
III.D. State versus National Recessions and Recoveries

One possible explanation behind the more disparate pace of recovery across states in the 1940s and 1950s might simply be that some states never entered a recession and, as a result, their unemployment rates remained relatively flat during national recoveries or even began rising. With a flat unemployment rate, equation (5) would estimate a very slow recovery rate during the national recovery, while a rising unemployment rate would generate a negative recovery rate. A greater share of states being relatively out of sync with the national business cycle in the 1940s and 1950s could thus generate the earlier dispersion of recovery rates depicted in figure 7. To explore this question, we first use our state recession dates to compute the share of states determined to be in a recession in each month and study how this share varies across the national business cycle.

Figure 8 plots the share of states identified as currently in a recession every month along with national NBER recession dates (gray bars).[41] The peak share of states experiencing a recession is roughly the same across the pre-1960s recessions and subsequent recessions, underscoring that the early postwar dispersion of recovery rates seen in figure 7 is not simply a product of many states not experiencing recessions during the national recessions of the 1940s–1950s. Another notable feature of figure 8 is that a number of national recovery periods show numerous states have yet to see unemployment recover. In particular, a sizable share of states remain coded as still experiencing a recession throughout the national business cycle expansions of 1954–1957, 1958–1960, 1970–1973, and 2001–2007. And a far greater share of states remain coded as in a recession in the two years following the 1990–1991 and 2001 recessions than any of the other NBER recessions, consistent with the particularly slow, uniformly paced state-level recoveries following these more recent downturns, as seen in figure 7. Broadly speaking, national recessions appear to be experienced

41. Though it is hard to see visually in the figure, 100 percent of states are being coded as "in recession" during the COVID-19 pandemic recession. The secondary rise in the share of states being coded as in recession after the initial onset of the pandemic occurs during the severe third national COVID-19 wave from November 2020–February 2021.

Figure 8. Share of US States in Recession, 1948–2022



Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS.

Note: Data sample from February 1948 to May 2022. State-level recession dates are estimated from the fitted claims-based unemployment rate for each state using the Dupraz, Nakamura, and Steinsson (2024) algorithm. The algorithm parameter is adjusted for each state proportionate to its average level of unemployment over the entire time period; see online appendix A.5 for details. Due to data limitations in nonfarm payroll employment, not all states are included early in this sample but are added when feasible; see footnote 7 for details.

more uniformly across states as recessions than do US business cycle expansions—particularly so in the first half of the postwar sample.[42]

A final concern we investigate is whether the early postwar dispersion in recovery rates depicted in figure 7 is simply an artifact of using uniform US recession dates to calculate state recovery rates, high and stable share of states in recession during US recessions notwithstanding. We instead analyze recovery paces normalized to state-specific business cycle troughs that occur near the end of national recessions. We match state business cycles to national business cycles by limiting our focus to state trough dates identified

42. Using our fitted series for state-level recession dating might be misleading if the inclusion of the US unemployment rate as a regressor in equation (2) causes state unemployment rates to track the national rate too closely in the pre-1976 out-of-sample predictions. As a robustness check, we also construct a version of figure 8 that instead uses our unfitted claims-based unemployment rates. These alternative state recession shares, which are plotted in online appendix figure B.11, are broadly consistent with those shown in figure 8.

by the Dupraz, Nakamura, and Steinsson (2024) algorithm within a plus or minus twelve-month window around the national trough dates identified from the official US unemployment rate, as reported in table 2.[43] In the case of double-dip recessions identified within each twelve-month window, we only measure the change in unemployment from the second trough.

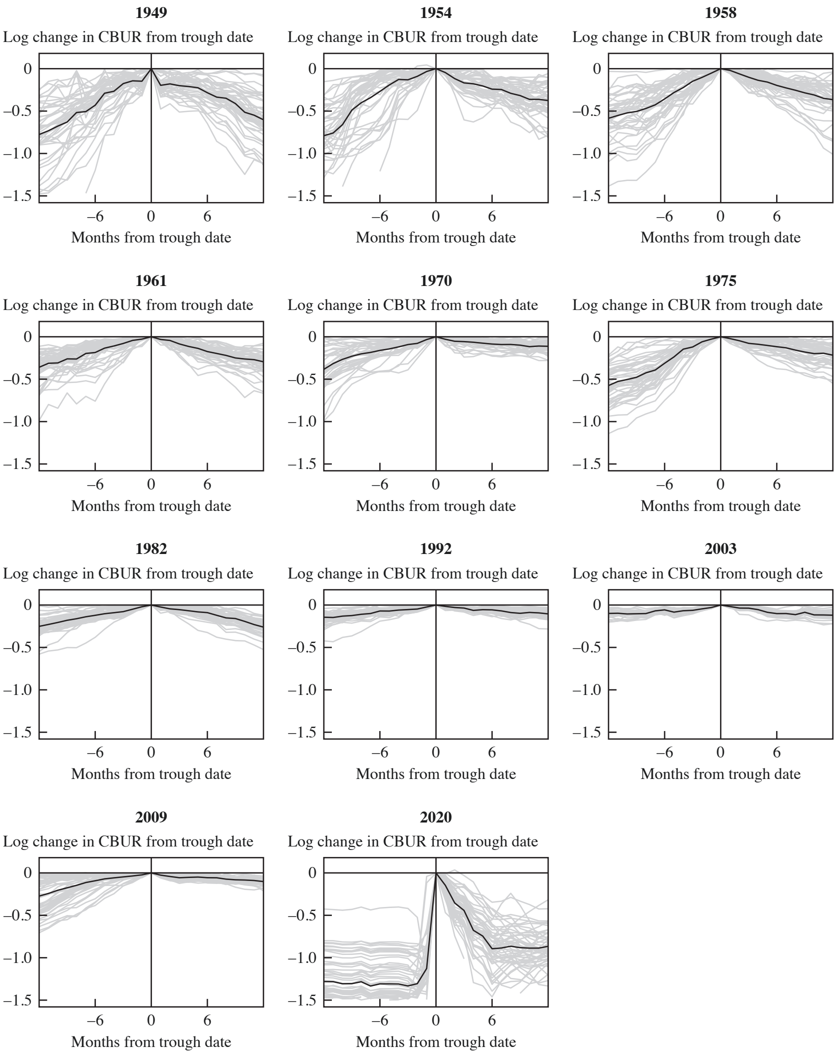
Figure 9 depicts the cumulative log point change in each state's fitted claims-based unemployment rate (gray lines) relative to that state's business cycle trough date. Figure 9 shows that the headline results depicted in figure 7 are robust to the choice of state or national recession dates: indexing to state-specific trough dates, there is again far more dispersion in the pace of states' unemployment recoveries following the 1948–1949, 1953–1954, and 1957–1958 recessions than during the recoveries from subsequent recessions—until the heterogeneity in recovery rates comes roaring back during the pandemic. And as in figure 7, albeit more visible here, figure 9 shows a more moderate degree of dispersion in state recovery rates following the 1960–1961, 1969–1970, 1973–1975, and 1981–1982 recessions, followed by much more consistent, slower paces of recovery following the 1990–1991, 2001, and 2007–2009 recessions. Pandemic recession aside, figure 9 again appears to reflect the emergence of a national business cycle that is increasingly experienced uniformly across almost all US states in the late twentieth century, mirroring the convergence of relative unemployment rates in figure 3 and states' unemployment recovery rates in figure 7.

Figure 9 also visually highlights the asymmetric speed with which our claims-based unemployment rates rise and fall, which is again consistent with dynamics of official unemployment rates. The bold black lines show the (unweighted) averages of the state-level changes in log unemployment, which underscore that unemployment tends to rise much faster during recessions than it falls during ensuing expansions, as seen for the US unemployment rate and claims-based unemployment rate in figure 1.[44] This is an important empirical regularity at both the state and national level that should be reflected in models of the business cycle, as Dupraz, Nakamura, and Steinsson (2024) emphasize at the national level; in particular, these

43. We marginally relax this window to thirteen months for the Great Recession, as an unusually large share of states (thirteen out of fifty) are identified as experiencing troughs exactly thirteen months from the national trough of October 2009; the latter is an unusual case of a relatively flat trough, and we thought it imprudent to throw away a quarter of those observations.

44. This tendency did, however, dissipate during the Great Moderation before reemerging in the Great Recession.

Figure 9. Unemployment Recoveries from States' Peak Unemployment Rate



Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS.

Note: Unemployment recoveries are measured as the cumulative log point change in each state's fitted claims-based unemployment rate relative to that state's peak unemployment rate (normalized to $t = 0$). Peaks in the fitted claims-based unemployment rates are identified within plus or minus twelve months of the peak US unemployment rate for each recession. In the case of double-dip recessions identified within the twelve-month window, we measure the cumulative change from the second peak in unemployment. Bold black lines are unweighted averages across states for each recession.

asymmetries favor unemployment models where recessionary shocks have a lasting impact on workers' employment outcomes.[45]

Lastly, figure 9 shows that recessions with greater dispersion in recovery rates are also clearly led by more varied rates of unemployment rising across states; the early postwar heterogeneity in recovery rates seen in figure 7 may thus partly be driven by greater variation in the magnitudes of shocks to unemployment across states. Consistent with this observation, the (pre-pandemic) standard deviations of relative unemployment rates reported in table 1 are highest around the US recessions of 1948–1949 and 1957–1958, which were followed by unusually fast, varied recovery rates. The cross-state convergence in the amplitude of unemployment rising during recessions and then falling during expansions, particularly since the late 1950s but also since the late 1970s, seems potentially important to the emergence of a more uniformly experienced postwar business cycle.

To cleanly study this evolving dynamic, we calculate state-specific cumulative changes in unemployment across postwar recessions and expansions, following Dupraz, Nakamura, and Steinsson (2024) but at the state level, using our state business cycle dates for the analysis.[46] Dupraz, Nakamura, and Steinsson (2024) document that the rise in the US unemployment rate during recessions is highly correlated with the decrease during the subsequent expansion, whereas the decrease in unemployment during expansions is uncorrelated with the rise in unemployment during the following recession; this asymmetry in the amplitude of unemployment supports the "plucking model" of business cycles of Friedman (1993), in which cyclical shocks pull output below potential, but the magnitude of shocks is unrelated to the strength of preceding expansions.

Panel A of figure 10 plots the amplitude of states' claims-based unemployment rates rising in contractions (x-axis) against the fall in unemployment during the ensuing expansion (y-axis); panel B plots the amplitude of states' unemployment falling during expansions against the rise during the ensuing contraction. Based on the evidence on the convergence in amplitude dynamics from figure 9, we differentiate between data before and after January 1980; data points up through 1979 are plotted as crosses and data for 1980 onward are plotted in circles.

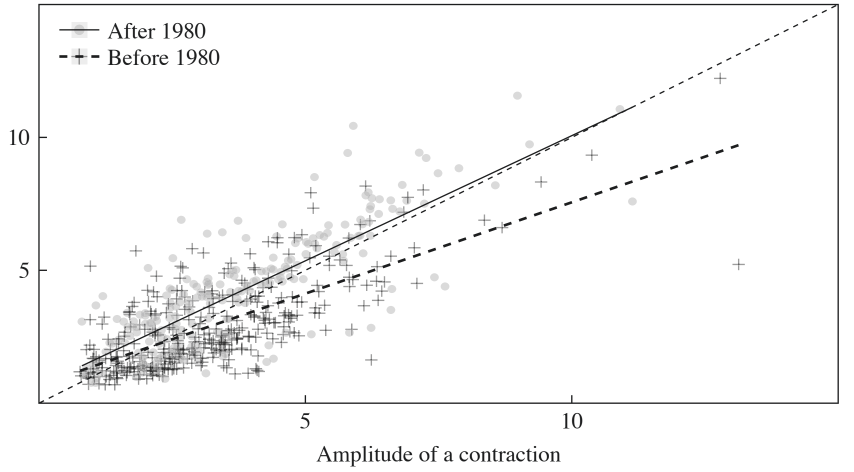
45. Gorry, Munro, and vom Lehn (2020) develop an alternative labor search model that yields such shock propagation.

46. We document that our claims-based unemployment rates emit the same amplitude dynamics at the national level; see online appendix B.4. Tasci and Zevanove (2019) also study these dynamics at the state level for 1976–2018, a sample limited by unemployment data availability; we find similar results over this more recent sample.

Figure 10. Amplitude of State-Level Claims-Based Unemployment Rates

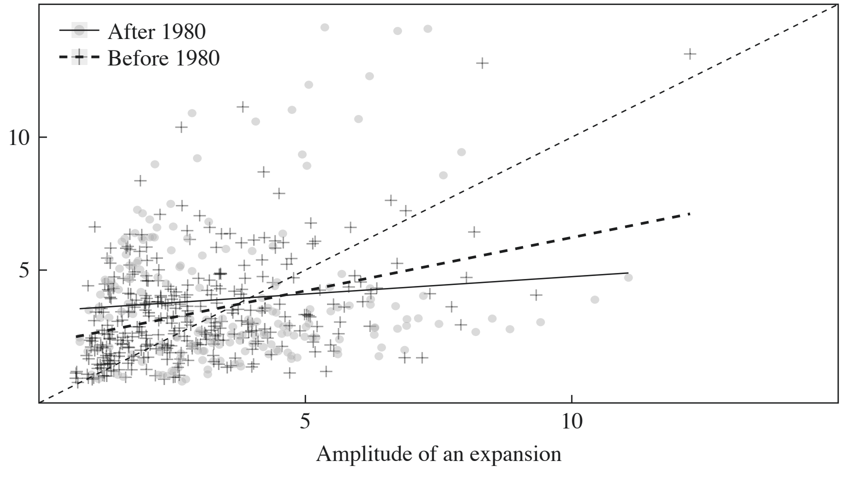
Panel A

Amplitude of the subsequent expansion



Panel B

Amplitude of the subsequent contraction



Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, and BLS.

Note: Data samples from January 1948 to December 2019. The amplitudes of contractions and expansions are measured as the percentage point change in our fitted claims-based unemployment rates between state peak and trough dates. We identify state-level recession dates using the Dupraz, Nakamura, and Steinsson (2024) algorithm on state-level claims-based unemployment rates. Pre-1980 data points are plotted in crosses (along with a corresponding dashed OLS regression line) and post-1979 data are plotted in circles (solid regression line). Finer dashed gray lines are 45-degree lines.

Panel A of figure 10 shows that our state-level claims-based unemployment rates also exhibit a strong positive correlation between the amplitude of unemployment rising during contractions and falling in subsequent recoveries; the association is significant at the 1 percent level for both the pre- and post-1980 samples. But this relationship has strengthened across states since 1980, with the correlation coefficient approaching unity: states are increasingly experiencing more uniform degrees of recovery in recent decades. In other words, they completely recover during expansions regardless of how much unemployment rose in the previous recession.[47] The flatter amplitude relationship in the pre-1980 data shows that unemployment not only tended to rise faster than it recovered, as seen in figure 9, but that more states had not fully recovered when their next recession hit; figure 8 similarly showed a higher average share of states coded as being in recession during US expansions before 1980 than after. The increasingly uniform degree, not just speed, of states' unemployment recoveries also points to the emergence of a national business cycle experienced more uniformly across states and may help explain the diminishing role of interstate migration as a margin of adjustment to adverse shocks since the early 1980s.

Panel B of figure 10 shows that our claims-based unemployment rates exhibit an insignificant correlation between the amplitude of unemployment in expansions and subsequent contractions for the post-1980 sample, but a positive correlation in the pre-1980 sample—one that is significant at the 1 percent level. Put differently, our historical state-level data show that the evidence in favor of the plucking model has actually strengthened since the late 1970s. Through the lens of the microfounded plucking model in Dupraz, Nakamura, and Steinsson (2024), which incorporates downward nominal wage rigidities into a search model to generate comparable amplitude dynamics, this implies that welfare gains from stabilization policy have, if anything, risen since the 1970s.

IV. The Emergence of a Uniform National Business Cycle

We have documented, over the postwar era: (1) a trend decrease in the standard deviation of relative employment growth and unemployment across states, both during recessions and expansions; (2) an attenuation of

47. The slope of the regression fit increases from 0.69 in the pre-1980 sample to 0.94 in the post-1980 sample; these correlation coefficients reflect the degree to which states' unemployment rates had, on average, recovered from the previous recession when the next recession hit.

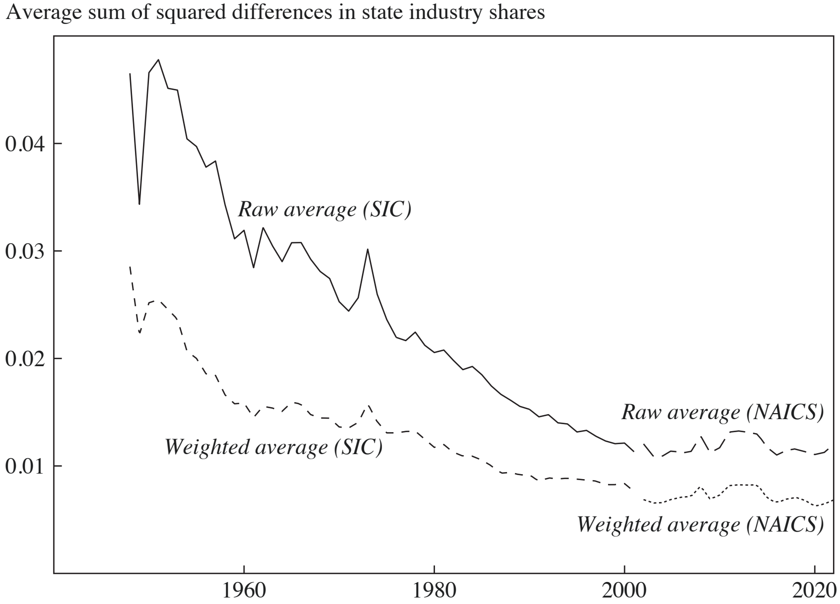
relative employment, unemployment, and population responses to state-specific demand shocks since the mid-1980s; and (3) convergence across states in both the speed and degree to which unemployment recovers after recessions. What can explain these related trends all pointing toward the emergence of a national business cycle experienced more uniformly across US states?

Our state unemployment recovery rates enable us to explore other state-level features that might be influencing the emergence of a more uniform business cycle, particularly the deceleration and convergence in unemployment recoveries after the late 1950s. To understand the convergence in unemployment recoveries first requires finding state-level features that are correlated with recovery paces in the early postwar recessions. Second, to account for the convergence over time, it must also be the case that those features have become more similar across states. We start by examining three key features that might be responsible for the convergence of business cycles across states over the postwar era: the convergence in industrial composition across states, the rise and convergence in female labor force participation rates (LFPR) across states, and the convergence in income per capita across states, as poorer regions caught up to richer ones.

We first document that states' industrial compositions have indeed become more similar over the postwar era, which could explain a convergence in both the speed and degree to which unemployment recovers after recessions if it means states experience shocks more similarly. Moreover, states sharing a more common industrial composition in more recent decades could mean fewer better economic opportunities in other industries elsewhere, inviting less out-of-state migration and explaining the diminished response of relative population to local shocks in recent decades. Similar to the construction of our Bartik instrument above, we construct annual national industry shares of earnings for the nine industries consistently available in the REA and compute the same industry shares for every state in each year. We calculate the sum of squared difference in these state-year industry shares relative to the nation-year industry shares and then average these differences across states. Figure 11 plots these annual averages—both unweighted and weighted by total earnings in each state—showing the evolving dispersion in industrial composition across states.

Figure 11 depicts a relatively rapid convergence in industrial composition from the start of our sample in the late 1940s to 1960, followed by a more gradual convergence from around 1960 to 2000; the degree of dispersion has been relatively narrow and stable since 2000. Notably, there was much greater dispersion in industrial composition across states during the faster,

Figure 11. Differences in State-Level Industrial Composition



Source: Authors' calculations based on BEA REA data.
Note: Data sample from 1948 to 2022. Figure plots the sum of squared differences in state-level industry shares relative to national industry shares, noting SIC or NAICS classifications.

disparate recoveries from the early postwar recessions in the 1940s–1950s. Over our sample of study, a big part of this industrial convergence was the transition of the US economy from manufacturing toward services and a knowledge economy.[48] As a simple proxy for exposure to the postwar convergence in industrial composition, we study states' manufacturing share of output.

States with particularly low female LFPR at the start of the gender revolution saw, on average, faster growth in female employment rates in subsequent decades (Fukui, Nakamura, and Steinsson 2023); this dynamic could be contributing to faster and more varied recovery rates earlier in our sample and more uniform, slower recovery rates after female LFPR stabilized around higher, more similar rates across states. To gauge any relationship

48. Manufacturing accounts for roughly 8 percent of employment since the Great Recession, down from about 25 percent in 1948.

between the convergences in female LFPR and unemployment recovery rates, we compute state-level female LFPR in 1950 using data from Fogli and Veldkamp (2011) as a proxy for differential exposure to this labor market shock, similar to the identification strategy used by Fukui, Nakamura, and Steinsson (2023). If there is an association between state-level recovery rates and female LFPR in the early postwar recessions, and these differences have diminished over time, this could be a mechanism driving more uniform unemployment dynamics across states in more recent decades.

Lastly, if poorer regions of the country had faster growth rates coming out of World War II, as they caught up to richer regions, this dynamic could also be contributing to faster and more varied unemployment recovery rates earlier in our sample of study; the gradually diminished persistence of employment growth seen in figure 4 could reflect such catch-up. To gauge any relationship between convergence in regional income and unemployment recovery rates, we construct measures of relative income for each state—computed as a state's per capita income divided by national per capita income—using REA data. Relative incomes have indeed converged across states: computing state-level Gini coefficients using per capita earnings (weighted by state population) we see a downward trend, with Gini coefficients falling from 0.12 in 1950 to 0.09 in 2000. If early postwar recovery rates are correlated with states' relative income and relative incomes have converged over time, this could again be contributing to increasingly uniform unemployment dynamics.

To explore any relationship between these trends of convergence, we run a simple multivariate regression, regressing states' average recovery pace in the first three "rapid" recoveries (1948–1958) on their manufacturing share of output (averaged over 1948–1957), female LFPR (in 1950), and relative income (averaged over 1950–1960). This empirical exercise is a simple way of gauging which state-level factors seem relevant to changing unemployment recovery rates and thus the emergence of a more uniform business cycle across states. Regression results are reported in table 3.

Of these trends of convergence we examine, table 3 shows that states' manufacturing share of output accounts for the largest share of the variance in unemployment recovery rates.[49] We find that female LFPR has no significant correlation with recovery paces in these early recessions.[50]

49. The adjusted R^2 of the regression in table 3 is 0.52 versus an adjusted R^2 of 0.49 in a univariate regression with only average manufacturing shares on the right-hand side.

50. Alternatively using the change in female LFPR from 1950 to 2000 also results in an insignificant correlation.

Table 3. Recovery Rate Regressions

Recovery pace	*Coeff.*	*Std. err.*	*p-value*
Manufacturing share	0.745***	0.132	0.000
Female LFPR	–0.004	0.005	0.374
Relative income	0.232**	0.088	0.012

Source: Authors' calculations based on data from the DOL, SSA, BLS, BEA, and Fogli and Veldkamp (2011).

Note: The dependent variable is the average state-level recovery rate over the national recoveries of 1949, 1954, and 1958. *, **, and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

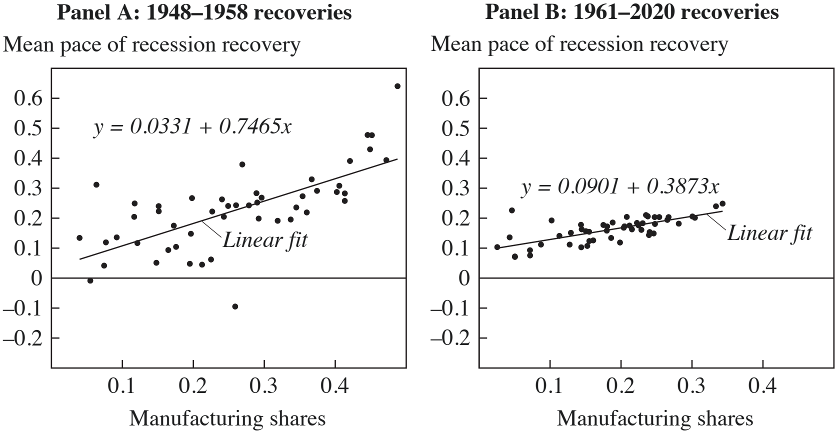
Lastly, we find states' relative income to be significantly correlated with recovery paces, though contrary to the catch-up dynamic we had in mind, relatively poorer states tended to recover more slowly after World War II. The explanatory power of relative incomes is considerably smaller than manufacturing shares, but the relationship is statistically significant at the 95 percent level. Thus, our results here suggest two partial explanations for states starting to experience recessions and recoveries more similarly: industrial composition is an important determinant of recovery rates and states' industrial compositions have become more uniform over time, while states with higher relative incomes tended to recover faster, but such income disparities across states have also diminished.

Because the convergence of industrial composition has by far the most explanatory power for early postwar unemployment recovery rates, we explore how this relationship in particular has evolved. Figure 12 plots states' average manufacturing share against their pace of economic recovery for two time periods: the first three rapid recoveries (1948–1958) and the subsequent seven slower recoveries (1961–2020) in panels A and B, respectively. Both samples show a positive, statistically significant correlation: states with larger manufacturing industries tend to experience more rapid recoveries in unemployment throughout the postwar era. The strength of this relationship diminishes substantially in the latter period but remains positive and significant.[51]

There could be multiple mechanisms behind states with larger manufacturing industries experiencing faster recoveries. Manufacturing-intensive states might be more adversely affected by recessions, generating a higher UR_0 in equation (5) and thus faster recoveries. Nearly every state is in a

51. The economic significance is somewhat weaker in the latter sample, but the linear relationships for both time periods are significant at the 1 percent level.

Figure 12. State-Level Recovery Rates by Manufacturing Share of Output



Source: Authors' calculations based on digitized and publicly available data from the DOL, SSA, BLS, and BEA.

Note: Data sample from 1948 to 1958 (panel A) and 1961 to 2020 (panel B). Recovery dates are estimated from the official US unemployment rate using the Dupraz, Nakamura, and Steinsson (2024) algorithm; see table 2. Recovery from the pandemic recession and the 1980 recession are excluded in panel B; see note for figure 6.

recession during most downturns, so for this explanation to hold, it must be that recessions in manufacturing-intensive states are more severe. Owyang, Piger, and Wall (2005) find that states with a higher manufacturing share of employment, particularly those in the Great Lakes region, contract relatively faster during recessions in their sample of study over 1979–2002, but our very negative Bartik shocks constructed over the full postwar sample show no such evidence of manufacturing-intensive states accounting for a disproportionate share of large negative shocks. Another possible mechanism is that the pace of recovery is affected by unique features of the manufacturing industry, for example, higher rates of unionization or the more intensive use of temporary layoffs; a national shift away from manufacturing could thus partly explain a slowdown in US recoveries.[52] Regarding the convergence in states' recovery rates, panel B of figure 12 also highlights that being a manufacturing-intensive state confers

52. See Lilien (1980) for evidence of high temporary layoff rates in the manufacturing sector; see Nekoei and Weber (2015) for evidence that temporary layoffs experience shorter unemployment spells; and see Gorry, Munro, and vom Lehn (2020) for a theoretical discussion about the importance of permanent displacements for the propagation of unemployment shocks.

less benefit in terms of faster recoveries in the post-1960 era; this could be a function of changes within manufacturing, such as deunionization, decreased reliance on temporary layoffs, and more permanent displacements during downturns. These questions are worthy of further examination using our newly constructed data set, but such an analysis is beyond the scope of this paper.

Many factors beyond the convergence in industrial composition and income per capita could also be contributing to the postwar emergence of a US business cycle experienced more uniformly across states. For one, significant changes in the degree of interstate economic integration since World War II could partly explain the deceleration and convergence in unemployment recoveries documented above. The abrupt decrease in both the dispersion and average pace of unemployment recoveries after the 1950s depicted in figure 7 occurs shortly after construction began on the Dwight D. Eisenhower National System of Interstate and Defense Highways, following the Federal-Aid Highway Act of 1956. The United States has also become far more integrated financially, particularly following widespread interstate and intrastate banking deregulation in the late 1970s through the early 1990s (Mian, Sufi, and Verner 2020); the timing of this shift toward a more deregulated, centralized, and nationwide banking system also lines up with the second convergence and deceleration in recovery rates between the 1973–1975 and 1990–1991 recessions, as seen in figure 9. Fiscal federalism has also shifted toward greater economic integration across states, particularly since the late 1960s.

Related to such increases in integration, our findings are consistent with general features of economic network models: shocks spread through the system more readily if the nodes of the network have higher connectedness (Kali and Reyes 2010; Giroud and Mueller 2019), and shocks with a more severe impact to more nodes in a network tend to be more severe for the network as a whole (Jackson 2010). Historical data availability makes it more challenging to quantify the potential relevance of these alternative mechanisms, but examining how the evolving network structure of the US economy relates to the convergence of state-level recovery rates is an exciting direction for future research.

While tradable goods and services markets have become far more integrated, it is well documented that interstate migration has fallen in recent decades. But US migration dynamics cannot seem to explain the timing of the sharp deceleration in recovery rates between the 1950s and 1960s; interstate migration rose rapidly following the Great Depression and then plateaued throughout the 1950s–1970s before starting to fall in the 1980s

(Rosenbloom and Sundstrom 2004; Molloy, Smith, and Wozniak 2011). If anything, the decreased persistence of employment following local shocks and convergence of state recovery rates—most pronounced since the Great Moderation—instead likely helps explain the decrease in interstate migration: if local job prospects gradually recover and economic prospects are not much better elsewhere, why incur the costs of relocating to a new state for economic reasons after a bad local shock?[53]

Lastly, while we have documented numerous trends pointing toward the emergence of a national business cycle experienced more uniformly across US states, greater dispersion in state recovery rates abruptly reemerged following the pandemic. Does that undermine our narrative?

We think not. There was unquestionably far more variance in the magnitude of the pandemic shock across states than any other postwar recession, as underscored by the record spike in the dispersion of relative unemployment rates seen in figure 3 and the wide-ranging increases in log unemployment seen in figure 9. Some of that variation appears to again reflect states' differential exposure from industrial composition, this time via the collapse of in-person services. For instance, Nevada was poised for a particularly bad shock, given its heavy reliance on tourism and leisure and hospitality services; Las Vegas casinos were shut down between March and June 2020, and Nevada saw the highest spike in either its claims-based unemployment rate or official unemployment rate of any state, followed by one of the fastest recovery rates. The variable timing of states experiencing waves of COVID-19 cases and when (or if) states introduced lockdowns is also surely contributing to some of the heterogeneity in unemployment dynamics. Many postwar shocks should have an impact on most states at roughly the same time, for example, the 1973 oil embargo or the 2008 financial crisis, but the pandemic spread more slowly and variably throughout the country, partly influenced by prevailing temperatures; for instance, figure 8 shows that roughly one-third of US states are identified as experiencing a double-dip recession during the third wave of late 2020 and early 2021, which had a differential and staggered degree of regional spread. An outlier in so many respects, it is interesting but perhaps unsurprising that post-pandemic data show much more of a divergence between state and national business cycles than the otherwise clear pre-pandemic, post-1950s trend.

53. Saks and Wozniak (2011) find that procyclical intercounty migration patterns are fairly stable in pre- and post-1980 samples, but they study national measures of labor market slack. Other factors are also surely contributing to falling migration rates, for example, home prices (Olney and Thompson 2024).

V. Concluding Remarks

In this paper, we introduce a new state-level unemployment data set spanning 1947–2023, a data set constructed from historical UI claims data that we digitized from a series of primary sources and then merged with existing state-level data for 1971 onward. We construct an (unfitted) claims-based unemployment rate series going back to January 1947 and an alternative fitted unemployment series going back to January 1948, estimating state unemployment rates from a statistical model of the dynamics between our claims-based unemployment rates and official measures of unemployment. We show that both claims-based unemployment rate series capture similar state and national business cycle dynamics as official data sources throughout overlapping samples. As the official BLS state unemployment rates only begin in January 1976, our data set represents a sizable expansion of panel data availability for measuring labor market slack, offering practitioners nearly three additional decades of seasonally adjusted monthly state-level data.

Our claims-based unemployment series significantly expands the scope for studying the evolution of local labor market adjustments to state-specific demand shocks as well as the evolving relationship between state and national business cycles. With our data set in hand, we document, over the postwar era: (1) a downward trend in the dispersion of relative employment growth and unemployment across states, during recessions and expansions; (2) an attenuation of relative employment, unemployment, and population responses to state-specific demand shocks since the mid-1980s; and (3) a convergence in both the speed and degree to which state unemployment rates recover after recessions. We argue that these are all related dynamics pointing toward the evolution of a national business cycle experienced more uniformly across states, especially since the late 1950s and also since the late 1970s.

We contribute some preliminary evidence on why a more uniform national business cycle emerged when it did. In particular, we provide some suggestive evidence that the convergence in industrial composition across states—particularly as relates to manufacturing shares of output—and the convergence in relative income per capita across states are related to the deceleration and convergence in unemployment recovery rates that we document above. By no means is our analysis exhaustive. Using our data set to study the role of increased interconnectedness—be it transportation, financial, or fiscal—would be an exciting direction for future research. Similarly, our data set would allow for a more comprehensive analysis of the evolving network structure of labor markets across US states, which is beyond the scope of this paper. A related avenue for future research would be constructing backdated state coincident indexes using our historical claims-based unemployment rates.[54] More broadly, we hope our historical data set of claims-based unemployment rates, derivative state recession dates, and the underlying digitized unemployment claims data prove useful for a wide range of empirical macroeconomic work using state-level panel data.

Lastly, this evolution from more disparate state business cycles into a more uniform national business cycle in the late twentieth century has important implications for macroeconomic stabilization policy. The deceleration and convergence in states' unemployment recovery rates in more recent decades make an even stronger case today than in the early postwar era for macroeconomic stabilization policy at the federal level to help accelerate recoveries across states experiencing more similar degrees of labor market slack. That fewer states have been experiencing relatively large local labor market demand shocks—and that these larger shocks have become relatively smaller, on average, in more recent decades—also reinforces the view that federal-level policies might appropriately spur faster recoveries across states experiencing business cycles more uniformly aligned with the national cycle.

ACKNOWLEDGMENTS We thank our editors Jón Steinsson and Janice Eberly and discussants Steven Davis and Marianna Kudlyak for helpful comments and suggestions. We also thank seminar and conference participants at the ASSA 2023 Annual Meeting, Economic Statistics Centre of Excellence Conference on Economic Measurement 2023, Southern Economic Association 2022 Annual Meeting, Midwest Macroeconomics Conference, Liberal Arts Macroeconomics Conference, University of Alberta, Middlebury College, and Liberal Arts Macro-Labor group for constructive feedback. We are especially grateful to Erica Cooper of the Department of Labor for providing digital scans of historical reports. We also thank Marianna Kudlyak and Jeremy Piger for providing data and Christopher House, Andrea Foschi, Linda Tesar, and Christian Pröbsting for helpful comments. David Munro gratefully acknowledges funding received from MiddData during work on this project. The views expressed in this paper are the views of the authors only and do not necessarily reflect the views of the Federal Reserve Bank of Dallas, the Federal Reserve System, or the International Monetary Fund, its executive board, or its management. A previous draft was circulated as "A New Claims-Based Unemployment Dataset: Application to Postwar Recoveries across U.S. States."

54. Nonfarm payroll employment and wage and salary disbursements, two of the three other inputs used in the latent factor model of Crone and Clayton-Matthews (2005), are also available at the state level back to 1948:Q1.

References

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.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, Mich.: W. E. Upjohn Institute for Employment Research.

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity,* no. 1: 1–75.

Brown, Jason P. 2017. "Identifying State-Level Recessions." *Economic Review* 102, no. 1: 85–108.

Bry, Gerhard, and Charlotte Boschan. 1971. "Programmed Selection of Cyclical Turning Points." In *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs.* New York: National Bureau of Economic Research.

Chodorow-Reich, Gabriel. 2019. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *American Economic Journal: Economic Policy* 11, no. 2: 1–34.

Crone, Theodore M., and Alan Clayton-Matthews. 2005. "Consistent Economic Indexes for the 50 States." *Review of Economics and Statistics* 87, no. 4: 593–603.

Dao, Mai, Davide Furceri, and Prakash Loungani. 2017. "Regional Labor Market Adjustment in the United States: Trend and Cycle." *Review of Economics and Statistics* 99, no. 2: 243–57.

Davis, Steven J., Prakash Loungani, and Ramamohan Mahidhara. 1997. "Regional Labor Fluctuations: Oil Shocks, Military Spending, and Other Driving Forces." International Finance Discussion Papers. Washington: Board of Governors of the Federal Reserve System.

Dupraz, Stéphane, Emi Nakamura, and Jón Steinsson. 2024. "A Plucking Model of Business Cycles." Working Paper. https://eml.berkeley.edu/~enakamura/papers.html.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2015. "On the Importance of the Participation Margin for Labor Market Fluctuations." *Journal of Monetary Economics* 72: 64–82.

Fogli, Alessandra, and Laura Veldkamp. 2011. "Nature or Nurture? Learning and the Geography of Female Labor Force Participation." *Econometrica* 79, no. 4: 1103–38.

Fort, Teresa C., and Shawn D. Klimek. 2018. "The Effects of Industry Classification Changes on US Employment Composition." Working Paper CES-18-28. Washington: United States Census Bureau.

Friedman, Milton. 1993. "The 'Plucking Model' of Business Fluctuations Revisited." *Economic Inquiry* 31, no. 2: 171–77.

Fukui, Masao, Emi Nakamura, and Jón Steinsson. 2023. "Women, Wealth Effects, and Slow Recoveries." *American Economic Journal: Macroeconomics* 15, no. 1: 269–313.

Giroud, Xavier, and Holger M. Mueller. 2019. "Firms' Internal Networks and Local Economic Shocks." *American Economic Review* 109, no. 10: 3617–49.

Glandon, P. J., Ken Kuttner, Sandeep Mazumder, and Caleb Stroup. 2023. "Macroeconomic Research, Present and Past." *Journal of Economic Literature* 61, no. 3: 1088–126.

Gorry, Aspen, David Munro, and Christian vom Lehn. 2020. "Experience, Skill Composition, and the Persistence of Unemployment Fluctuations." *Labour Economics* 63: 101793.

Hall, Robert E., and Marianna Kudlyak. 2020. "The Inexorable Recoveries of US Unemployment." Working Paper 28111. Cambridge, Mass.: National Bureau of Economic Research.

Hall, Robert E., and Marianna Kudlyak. 2022. "The Inexorable Recoveries of Unemployment." *Journal of Monetary Economics* 131: 15–25.

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57, no. 2: 357–84.

Hamilton, James D. 2018. "Why You Should Never Use the Hodrick-Prescott Filter." *Review of Economics and Statistics* 100, no. 5: 831–43.

Hazell, Jonathon, Juan Herreño, Emi Nakamura, and Jón Steinsson. 2022. "The Slope of the Phillips Curve: Evidence from U.S. States." *Quarterly Journal of Economics* 137, no. 3: 1299–344.

Hodrick, Robert J., and Edward C. Prescott. 1997. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking* 29, no. 1: 1–16.

Jackson, Matthew O. 2010. *Social and Economic Networks*. Princeton, N.J.: Princeton University Press.

Kali, Raja, and Javier Reyes. 2010. "Financial Contagion on the International Trade Network." *Economic Inquiry* 48, no. 4: 1072–101.

Lilien, David M. 1980. "The Cyclical Pattern of Temporary Layoffs in United States Manufacturing." *Review of Economics and Statistics* 62, no. 1: 24–31.

McMurrer, Daniel P., and Amy B. Chasanov. 1995. "Trends in Unemployment Insurance Benefits." *Monthly Labor Review* 118: 30–39.

Mian, Atif, Amir Sufi, and Emil Verner. 2020. "How Does Credit Supply Expansion Affect the Real Economy? The Productive Capacity and Household Demand Channels." *Journal of Finance* 75, no. 2: 949–94.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic Perspectives* 25, no. 3: 173–96.

Nakamura, Emi, and Jón Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104, no. 3: 753–92.

Nekoei, Arash, and Andrea Weber. 2015. "Recall Expectations and Duration Dependence." *American Economic Review* 105, no. 5: 142–46.

Neumann, George R., and Robert H. Topel. 1991. "Employment Risk, Diversification, and Unemployment." *Quarterly Journal of Economics* 106, no. 4: 1341–65.

Olney, William W., and Owen Thompson. 2024. "The Determinants of Declining Internal Migration." Working Paper 32123. Cambridge, Mass.: National Bureau of Economic Research.

Owyang, Michael T., Jeremy Piger, and Howard J. Wall. 2005. “Business Cycle Phases in U.S. States.” *Review of Economics and Statistics* 87, no. 4: 604–16.

Romer, Christina D., and David H. Romer. 2019. “NBER Business Cycle Dating: Retrospect and Prospect.” Paper prepared for the ASSA Annual Meeting, San Diego, Calif., January 3–5, 2020.

Rosenbloom, Joshua L., and William A. Sundstrom. 2004. “The Decline and Rise of Interstate Migration in the United States: Evidence from the IPUMS, 1850–1990.” In *Research in Economic History* 22: 289–325. Leeds: Emerald Group Publishing Limited.

Sahm, Claudia. 2019. “Direct Stimulus Payments to Individuals.” In *Recession Ready: Fiscal Policies to Stabilize the American Economy,* edited by Heather Boushey, Ryan Nunn, and Jay Shambaugh. Washington: Brookings Institution Press.

Saks, Raven E., and Abigail Wozniak. 2011. “Labor Reallocation over the Business Cycle: New Evidence from Internal Migration.” *Journal of Labor Economics* 29, no. 4: 697–739.

Stock, James H., and Mark W. Watson. 2010. “The Evolution of National and Regional Factors in U.S. Housing Construction.” In *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, edited by Tim Bollerslev, Jeffrey Russell, and Mark W. Watson. Oxford: Oxford University Press.

Tasci, Murat, and Nicholas Zevanove. 2019. “Do Longer Expansions Lead to More Severe Recessions?” *Economic Commentary,* Federal Reserve Bank of Cleveland, January 14. https://www.clevelandfed.org/publications/economic-commentary/2019/ec-201902-do-longer-expansions-lead-to-more-severe-recessions.

Yagan, Danny. 2019. “Employment Hysteresis from the Great Recession.” *Journal of Political Economy* 127, no. 5: 2505–58.

Comments and Discussion

COMMENT BY

STEVEN J. DAVIS Fieldhouse, Munro, Koch, and Howard make two sets of useful contributions. First, they collect and digitize monthly state-level unemployment statistics that extend back to January 1947. Second, they develop new evidence about how the nature of US business cycle fluctuations has evolved since 1947. In what follows, I first make some remarks about prior research and the interpretation challenges that arise with the authors' claims-based unemployment rate. I then turn to the paper's central empirical finding—the emergence of greater uniformity across states in US labor market fluctuations—and draw out some related implications.

STATE-LEVEL UNEMPLOYMENT STUDIES: WHAT CAME BEFORE The authors skim lightly over previous studies that exploit state-level unemployment statistics before 1976. For more than sixty years, the Bureau of Labor Statistics and the Department of Labor (sometimes in conjunction with the Employment Training Administration or the Manpower Administration) have been releasing state and local area unemployment statistics that extend back to the 1950s or even earlier. There is an old literature on the quality of these measures and how to improve them. Examples include Ullman (1963) and Wetzel and Ziegler (1974).

Later studies exploited these state-level unemployment statistics. Neumann and Topel (1991), for example, analyze how the industry mix of employment in a state influences its average unemployment rate over time and its response to common aggregate disturbances. Their analysis covers the period from 1950 to 1985. Davis, Loungani, and Mahidhara (1997) study the effects of oil shocks, defense contract awards, military basing, and Bartik-like industry mix shocks on state-level unemployment

Brookings Papers on Economic Activity, Spring 2024: 320–341 © 2024 The Brookings Institution.

and employment growth fluctuations from 1956 to 1992. Our analysis of state-level fluctuations finds meaningful effects for each of these categories of shocks, a larger near-term effect of negative shocks than positive ones, the spillover of own-state labor demand shocks to nearby states, and a slow equilibration of state-level unemployment rates. Many studies use state-level unemployment data for the period from 1976 onward.

MEASUREMENT AND INTERPRETATION CHALLENGES As Corson and Nicholson (1988, 62) discuss and analyze at length, "the characteristics of state [unemployment insurance] UI programs . . . varied considerably" with respect to minimum qualifying wages, the eligibility of persons who quit their jobs, the stringency of prior work tests, claims denial practices, maximum weekly benefits, wage replacement rates, pension and Old-Age, Survivors, and Disability Insurance (OASDI) offset provisions in benefit determination, and employer-initiated challenges to claims for unemployment benefits. As they also explain, many characteristics of state-level UI programs changed in the 1970s and 1980s—differently so across states—in ways that affected the share of unemployed persons who claimed and collected unemployment benefits. Major changes to federal policy in the 1970s and 1980s also brought changes to state-level UI programs. While these federal policy developments induced roughly coincident changes at the state level in claims for unemployment benefits and in the ratio of insured to total unemployment, the extent of these state-level responses to federal policy changes differed across states because of states' differences in industry mix, demographic structure, and prior UI program characteristics.

Over time, federal and state policy changes led to more uniformity in UI program characteristics across states. Policy changes also contributed to an enormous expansion in the share of employment in industry sectors covered by the UI system. According to McMurrer and Chasanov (1995), civilian workers in sectors covered by the UI system rose from 56 percent in 1947 to 90 percent or more by the mid-1970s.[1] Expanding UI coverage also played out differently across states because of state-level differences in industry mix and coverage criteria before the 1980s. In short, UI program characteristics and coverage became more similar across states over time, a process that continued through at least the early 1980s.

These observations lead to the following conclusions. First, state-level fluctuations from 1947 to the early or mid-1980s in the claims-based unemployment rate, the insured unemployment rate, and the ratio of insured or

1. See McMurrer and Chasanov (1995), chart I.

claims-based unemployment to total unemployment are partly an artifact of: (a) state-specific changes to UI program characteristics; and (b) differential state-level effects of changes in federal policy. Second, the claims-based unemployment statistics overstate the drift to more similarity in state-level fluctuations from 1947 to the mid-1980s. The force of this point is greater for the claims-based unemployment rate than for the insured unemployment rate because the former is especially sensitive to changes in UI program coverage. Third, using state and time fixed effects to control for UI coverage and program characteristics does not fully address this overstatement in the trend toward greater cross-state uniformity in unemployment fluctuations from 1947 to the mid-1980s, because the state-level changes happened at different times in different states and because the federal changes had uneven effects across states.

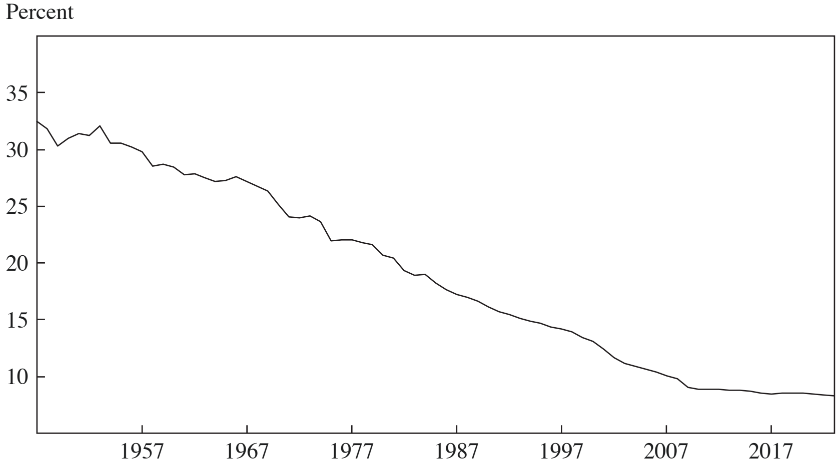
GREATER SIMILARITY ACROSS STATES IN LABOR MARKET FLUCTUATIONS Despite concerns about the claims-based unemployment measure, I am persuaded that there was a drift to less dispersion in unemployment rates across US states from 1947 through 2019, and that the severity and timing of recessions and recoveries became more similar across states. I am persuaded because the authors support these claims using other types of data, including state-level data on employment growth rates. They also show that, over time, the industry mix of economic activity became more similar across states, which helps explain "the emergence of a uniform business cycle."

These are important findings, in part because they reflect slow-working changes in the structure of the economy that are likely to influence the nature of future US business cycle fluctuations.[2] I say "likely" rather than "inevitable" because future shocks of an extraordinary nature could drive a sharp departure from prior trends, as with the COVID-19 recession and recovery.

WILL FUTURE LABOR MARKET TRANSFORMATIONS BRING LESS HARDSHIP AND DISLOCATION THAN DID DEINDUSTRIALIZATION? Conditional on the pace of transformation, there are good reasons to think so. The first reason follows directly from the now greater similarity across states in the industry mix of employment. Hence, looking forward, future US recessions are likely to be more evenly felt across locations than past recessions. As a result, future recessions

2. In a distinct finding, the authors provide evidence that the pace of unemployment rate declines after recessions has slowed over time. For the sake of brevity, I do not discuss that finding here. The interested reader may consult my slides on the topic, which accompanied my remarks at the Spring 2024 *BPEA* Conference. https://www.brookings.edu/events/bpea-spring-2024-conference/.

Figure 1. Manufacturing's Share of Nonfarm Employees, 1947–2023



Source: Bureau of Labor Statistics, series PAYEMS and MANEMP, retrieved from FRED; and author's calculations.

(and labor market transformations more broadly) are likely to require less geographic mobility of factor inputs. Geographic mobility is costly, and sometimes risky, not least for displaced workers and their families when compelled to move by job loss and locally depressed economies.

A second set of reasons follows from some distinctive characteristics of the manufacturing sector. Historically, manufacturing industries often played outsize roles in their local economies to an extent that was and is uncommon for other major sectors. In addition, most manufacturing industries are cyclically sensitive—again, more so than other major sectors except for construction.

Indeed, most US recessions since 1945 have involved large, spatially concentrated, contractions in the manufacturing sector. As a result, many manufacturing workers lost jobs in the same places at the same time. That made it harder for each job loser to find an attractive new job. Other things equal, losing a job in a recession has a much greater negative impact on future earnings than losing a job in an expansion (Davis and von Wachter 2011). Because job losses in the manufacturing sector were concentrated, spatially and temporally, they led to worse outcomes, on average, for individual job losers.

Manufacturing's share of employment has plummeted over the past seventy-five years (figure 1). As of 2023, manufacturing jobs account for only 8.3 percent of nonfarm payroll employees in the US economy. Thus, there is simply less scope for spatially and cyclically concentrated downturns in the manufacturing sector to drive high-volume job losses in the future than in the past.

Third, the biggest—or, at least, most hyped—prospective source of transformation on the horizon is the possible rollout of artificial intelligence (AI) technologies that automate jobs and displace workers. It's hard to predict whether and when AI technologies will displace large numbers of workers. However, the available evidence says that AI-exposed jobs are distributed across a broad range of industries (Acemoglu and others 2022). Most of these industries are less cyclically sensitive than manufacturing. Hence, future job losses in these industries—whether due to AI or other forces—won't be as concentrated in recessions as manufacturing job losses. And they won't be as spatially concentrated. Hence, for a given pace of job loss, future recessions and labor market transformations are likely to involve less economic hardship and dislocation than past manufacturing job losses.

Moreover, to the extent that AI-exposed workers and jobs are concentrated in particular states and areas, they tend to be concentrated in the more densely populated parts of the country (CEA 2024).[3] Thus, AI-driven job losses are also likely to be more concentrated in densely populated areas than the job losses associated with the deindustrialization of US employment. Other things equal, the larger and thicker labor markets found in dense urban centers lead to shorter nonemployment spells and smaller earnings losses for job losers (Moretti and Yi 2024).

A final set of reasons flows from the pandemic-instigated shift to work from home (Barrero, Bloom, and Davis 2021). By relaxing locational constraints, this development enlarges the geographic reach of labor markets for fully remote and hybrid jobs and for the workers who fill them. Effectively, the labor markets for these jobs are now larger and thicker, which is another reason to anticipate shorter nonemployment spells and smaller earnings losses for job losers in remote-suitable occupations.

The shift to remote work also moderates the economic hardship associated with job loss for a related reason: employer-level workforces are becoming more geographically dispersed. Akan and others (2024) provide some evidence on this score by exploiting data from Gusto, a firm that provides payroll processing and other services to mostly smaller and midsize employers. They analyze employee-level data linked to a balanced

3. See CEA (2024), figures 7–10.

panel of 5,800 firms that operated continuously from 2018 to 2023. As of 2019, less than 1 percent of the employees at these firms resided more than fifty miles from their employer's worksite. By 2023, about 2 percent of the employees hired before March 2020 at these firms lived more than fifty miles away. Among employees at these firms who were hired since March 2020, more than 7 percent live more than fifty miles away from the employer's worksite as of 2023. This result says that a partial untethering of worker residential locations from employer worksite locations is underway. That process will continue to unfold for many years as company workforces gradually turn over. That's yet another reason to anticipate that firm-wide (and industry-wide) contractions in the future will be less spatially concentrated.

REFERENCES FOR THE DAVIS COMMENT

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. 2022. "Artificial Intelligence and Jobs: Evidence from Online Vacancies." *Journal of Labor Economics* 40, no. S1: S293–S340.

Akan, Mert, Jose Maria Barrero, Nicholas Bloom, Tom Bowen, Shelby Buckman, Steven J. Davis, Luke Pardue, and Liz Wilke. 2024. "Americans Now Live Farther from Their Employers." Stanford, Calif.: Hoover Institution.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2021. "Why Working from Home Will Stick." Working Paper 28731. Cambridge, Mass.: National Bureau of Economic Research.

Corson, Walter, and Walter Nicholson. 1988. *An Examination of Declining UI Claims during the 1980s: Draft Final Report.* Princeton, N.J.: Mathematica Policy Research.

Council of Economic Advisers (CEA). 2024. *Economic Report of the President*. Washington: White House. https://www.whitehouse.gov/wp-content/uploads/2024/03/ERP-2024.pdf.

Davis, Steven J., Prakash Loungani, and Ramamohan Mahidhara. 1997. "Regional Labor Fluctuations: Oil Shocks, Military Spending, and Other Driving Forces." International Finance Discussion Papers 578. Washington: Board of Governors of the Federal Reserve System.

Davis, Steven J., and Till von Wachter. 2011. "Recessions and the Costs of Job Loss." *Brookings Papers on Economic Activity*, Fall: 1–55.

McMurrer, Daniel P., and Amy B. Chasanov. 1995. "Trends in Unemployment Insurance Benefits." *Monthly Labor Review* 118: 30–39.

Moretti, Enrico, and Moises Yi. 2024. "Size Matters: Matching Externalities and the Advantages of Large Labor Markets." Working Paper 32250. Cambridge, Mass.: National Bureau of Economic Research.

Neumann, George R., and Robert H. Topel. 1991. "Employment Risk, Diversification, and Unemployment." *Quarterly Journal of Economics* 106, no. 4: 1341–65.

Ullman, Joseph C. 1963. "How Accurate Are Estimates of State and Local Unemployment?" *Industrial and Labor Relations Review* 16, no. 3: 434–52.

Wetzel, James R., and Martin Ziegler. 1974. "Measuring Unemployment in States and Local Areas." *Monthly Labor Review* 97, no. 6: 40–46.

COMMENT BY

MARIANNA KUDLYAK Fieldhouse, Munro, Koch, and Howard construct historical monthly unemployment series for US states dating back to January 1947 using newly digitized unemployment insurance claims data. Their analysis reveals an attenuation of relative employment, unemployment, and population responses to state-specific demand shocks in recent decades. They document the emergence of a national business cycle, characterized by state unemployment rates recovering at speeds similar to the aggregate unemployment rate, averaging 10 percent per year. Additionally, they observe a convergence in the industrial composition across states.

The authors introduce new data series and present a rich set of results on regional evolutions and unemployment recoveries. These series significantly enhance economists' ability to conduct both regional and aggregate-level labor market and business cycle analysis, providing a valuable service to the profession. My comments focus on three areas: the conceptual differences between claims-based and official unemployment measures, the interpretation of results on regional evolutions, and the unemployment recovery from the pandemic recession.

CLAIMS-BASED VERSUS OFFICIAL UNEMPLOYMENT AS MEASURES OF LABOR MARKET SLACK I would like to start by focusing on the differences and similarities between the new claims-based unemployment rate versus the official unemployment rate as measures of labor market slack. I argue that the two measures have conceptual dissimilarities. The question then remains whether the two measures provide the same information about the timing and the degree of labor market slack, as well as how the measures relate to other variables over the business cycle.

The claims-based unemployment rate is a conceptually different measure of labor market slack as compared to the official unemployment rate. Specifically, the official unemployment rate counts all the nonemployed who self-report actively searching for work or those nonemployed who are not searching but are on temporary layoff. The claims-based measure covers those nonemployed who are eligible for and who claim their unemployment insurance (UI) benefits. State unemployment programs typically exclude certain workers from benefit eligibility, notably agricultural workers, domestic workers, and independent contractors; federal employees and veterans instead qualify for federally administered UI programs. More generally, eligibility depends primarily on three factors, all of which are determined at the state level. The three factors are pre-separation minimum duration of employment and pay (monetary criteria), the duration of unemployment, and voluntary versus nonvoluntary nature of the separation.

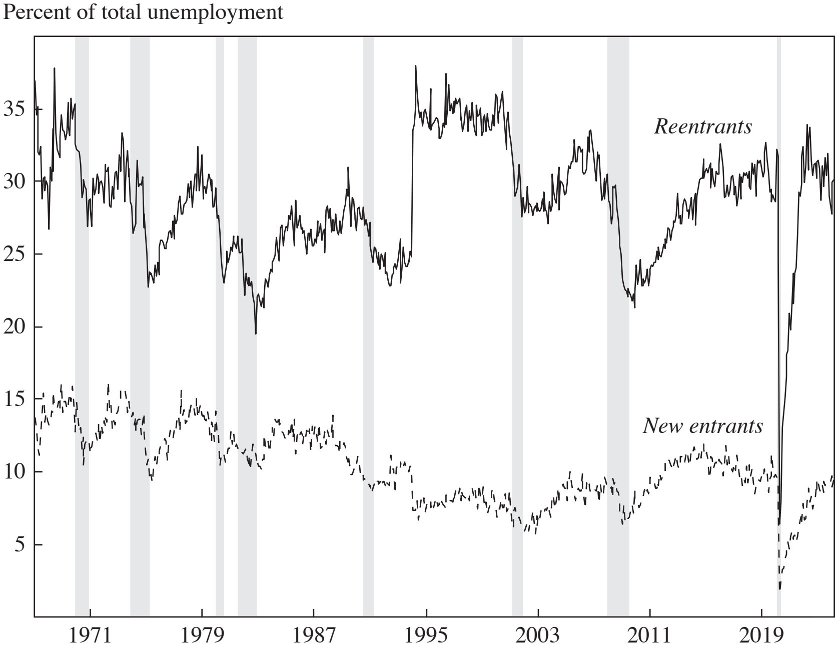
A significant portion of the unemployed under the official unemployment rate definition are not eligible for UI. These groups include: (1) individuals who hold short-term jobs; (2) labor force new entrants or reentrants; (3) quits; and (4) the long-term unemployed who exhausted their UI benefits.

The individuals who hold short-term jobs and circle between work and nonwork represent a large fraction of unemployment but are less likely to be eligible for UI. Eligibility requirements vary by state, but generally, there is a minimum requirement of work hours and employment duration for eligibility. As such, a claims-based measure might undercount those individuals. This might have implications for measuring labor market slack if the share of such individuals among all the unemployed varies over the business cycle. For example, Hall and Kudlyak (2019) find that most of unemployment comes from a small fraction of the population that tends to circle between employment, being out of the labor force, and short-term jobs. Gregory, Menzio, and Wiczer (2024) and Ahn, Hobijn and Şahin (2023) find that such individuals constitute a larger share of the unemployed during recessions than during recoveries.

The new entrants and reentrants in unemployment are typically ineligible for UI and therefore are not included in the UI claims data. Figure 1 shows that these series represent a procyclical share of total unemployment.

Another issue is the long-term unemployed and exhaustion of benefits. One aspect of eligibility for UI is that the individuals who exhausted their UI benefits—the long-term unemployed (LTU)—are not eligible for UI. These individuals are not counted in the claims-based measure. The share of LTU in unemployment is countercyclical (see figure 2). The issue is especially acute in the deep recession of 2007–2009. The authors calculate their claim-based measure from the regular state programs and exclude the extensions such as Federal-State Extended Benefits and Emergency Unemployment Compensation 2008. The authors motivate their choice by the idea of creating a consistent measure of labor market slack. However, the extended programs are designed to alleviate the labor market slack; therefore, not counting the individuals on such programs might undermeasure the labor market slack.

Figure 1. The New Entrants and Reentrants as Percentage of Total Unemployment



Source: Author's calculations using data from the Current Population Survey (CPS) and National Bureau of Economic Research (NBER), retrieved from Haver.

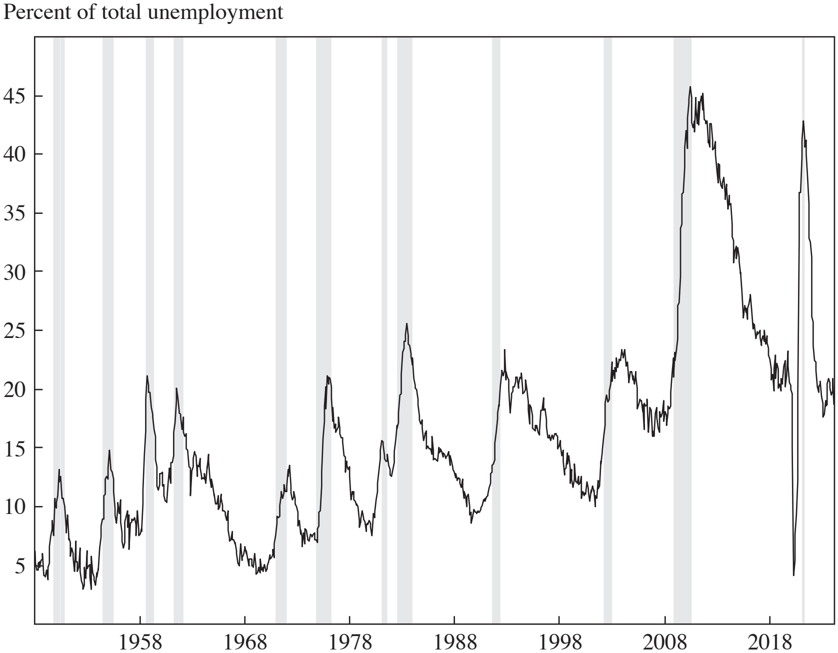
Additionally, the claim-based measure does not count quits. Figure 3 shows that quits represent a procyclical share of total unemployment.

Separate from UI eligibility is an issue of UI take-up rate. The UI take-up rate varies over time and by state. Auray, Fuller, and Lkhagvasuren (2019) find that from 1989 to 2012, on average, 23 percent of those eligible for UI benefits in the United States did not collect them. They show that the general pattern of the take-up rate over time shows an upward trend from 1989 to 2002, with a downward trend thereafter.[1] Furthermore, there is substantial variability of take-up rate by state—from 38 percent to 95 percent (figure 4).

The discussion above demonstrates that there are conceptual differences between the claims-based unemployment rate and the official unemployment rate. Figure 5 shows the two series (panel A) and the difference between

1. See Auray, Fuller, and Lkhagvasuren (2019), figure 1.

Figure 2. Long-Term Unemployment (Six Months or More) as Share of Total Unemployment



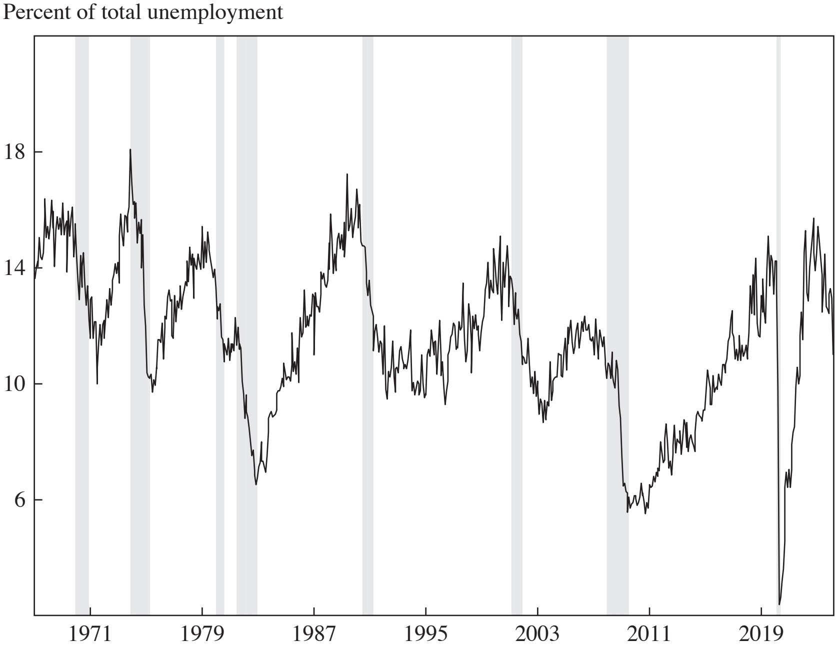
Source: Author's calculations using data from CPS and NBER, retrieved from Haver.

the official rate and the claims-based rate (panel B). Fieldhouse, Munro, Koch, and Howard document that their new state-level measures capture business cycle inflection points and amplitude dynamics in line with those captured by the official unemployment rate. Figure 5, however, shows that the two measures signal different degree of slack and that the difference between the measures varies over the business cycle in a systematic manner. Specifically, the difference appears countercyclical, and it is especially large when the conceptual differences described above are especially acute—for example, during the 2007–2009 recession.

One area of potential future research is an investigation of which of the two measures of the unemployment rate better correlates with other measures over the business cycle.

REGIONAL EVOLUTIONS REVISITED The authors use the newly constructed state unemployment rates to revisit one of the long-standing questions in economics—the adjustment of labor markets to local employment shocks. In a seminal work, Blanchard and Katz (1992) ask: when a typical US state

Figure 3. Quits as Share of Total Unemployment

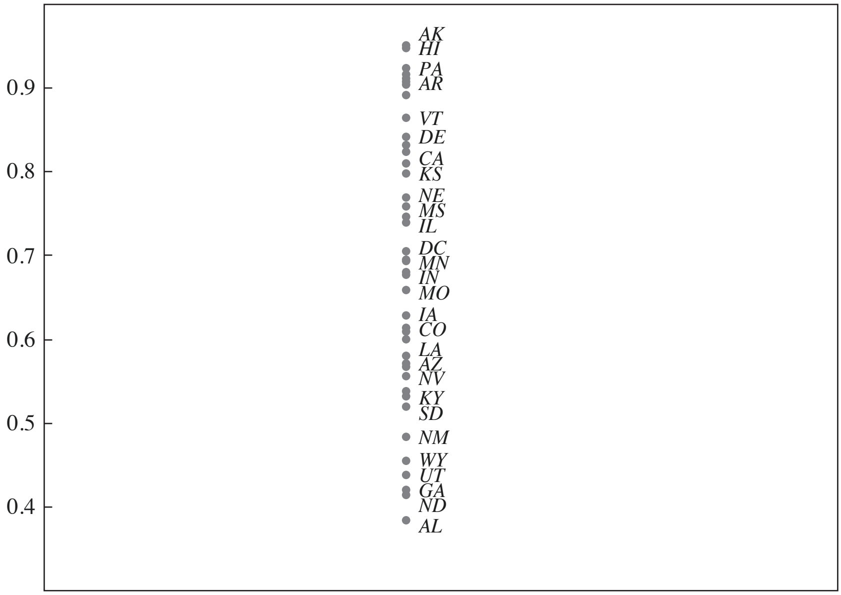


Source: Author's calculations using data from CPS and NBER, retrieved from Haver.

has been affected by an adverse shock to employment, how has it adjusted? Specifically, did wages decline relative to the rest of the nation? Were other jobs created to replace those jobs destroyed by the shock? Or did workers move out of the state? Blanchard and Katz (1992) develop a model of regional evolutions, which points to two adjustment mechanisms: first, lower wages and higher unemployment in response to a negative local employment shock induce net out-migration of labor; second, lower wages induce net in-migration of firms and the creation of jobs. The long-run effect on employment depends on the relative strength and speed of the two mechanisms.

Applying a vector autoregression (VAR) framework to the state data over 1978–1990, Blanchard and Katz (1992) find that relative employment responds strongly and remains persistently depressed while relative unemployment and labor force participation exhibit muted, transitory responses. Wages exhibit a transitory decline of about 0.50 percent, while house prices decline by 2 percent, implying a consumption wage decline of only 0.20 percent. Blanchard and Katz (1992) interpret the findings that

Figure 4. Unemployment Insurance Take-Up Rates across US States, Given Eligibility



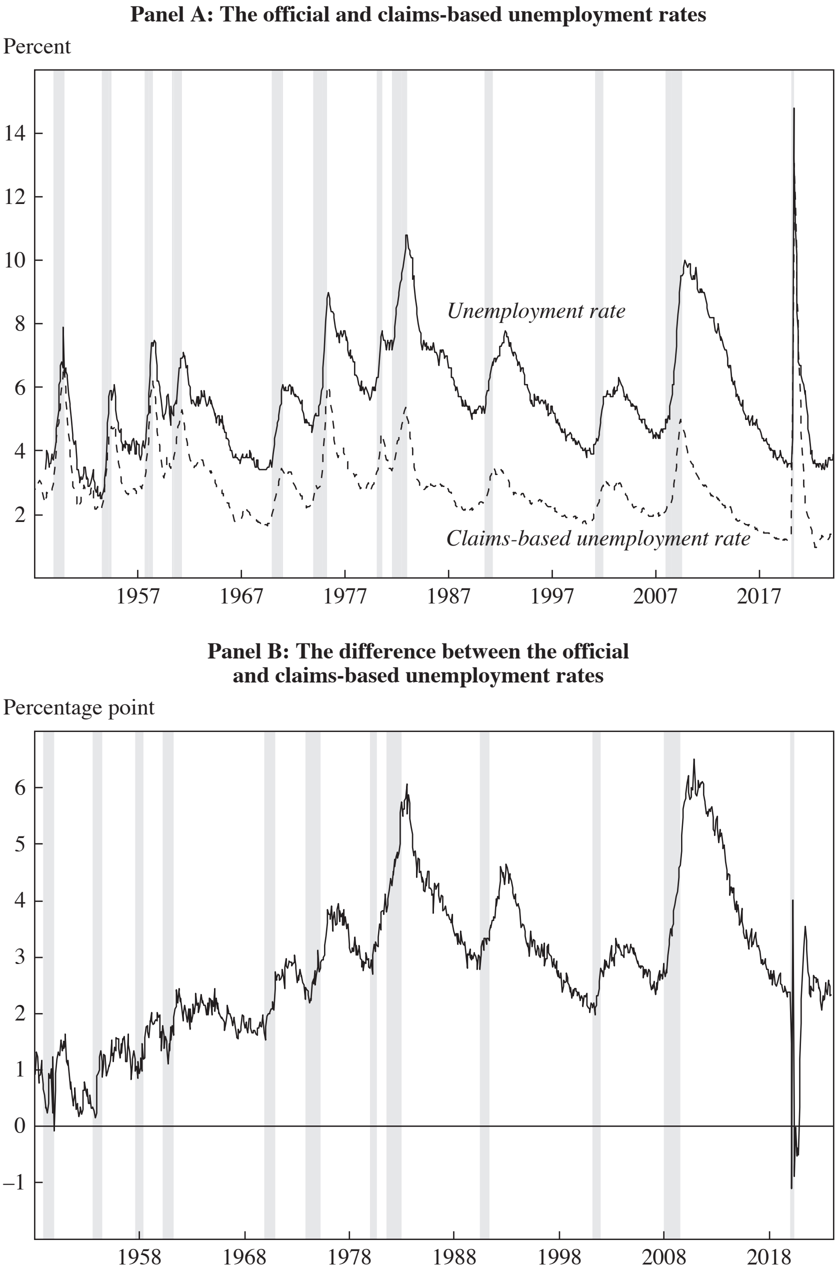
Source: Author's calculations using data from Auray, Fuller, and Lkhagvasuren (2019).

in the longer run, quantities, not prices, make the adjustment. There is relatively little shock absorption by wages. They find that the dominant adjustment mechanism is labor mobility rather than job creation. Labor mobility is a response to unemployment rather than consumption wages. That is, migration accounts for most of the adjustment to local labor demand shocks.

Fieldhouse, Munro, Koch, and Howard, using a local projections method with the Bartik shock and applying it to the newly constructed data over 1949–2019, find an attenuation of relative employment, unemployment, and population responses to state-specific demand shocks in recent decades. They conclude that the role of interstate migration has diminished and suggest that the states' increasingly similar industrial composition might help explain why interstate migration is becoming less of a margin for adjustment to local demand shocks.

The key question in interpreting the authors' estimates is whether the diminished estimates represent changes in responses to local shocks or changes in the outside options. Borusyak, Dix-Carneiro, and Kovak (2023) demonstrate that workers' reallocation decision depends both on the shock to their current location as well as the shocks to the potential alternative

Figure 5. The Difference between the Official Unemployment Rate and the Claims-Based Unemployment Rate



Source: Author's calculations using data from CPS and NBER (retrieved from Haver) and the paper by Fieldhouse, Munro, Koch, and Howard.

locations. Therefore, by omitting the shocks to the relevant alternative locations, the conventional migration regression (estimated by Fieldhouse, Munro, Koch, and Howard) is misspecified. Borusyak, Dix-Carneiro, and Kovak (2023) show that attenuation of the shock effect in such conventional regressions is particularly severe if the shocks are correlated across locations. Furthermore, they show that when labor demand shocks have an industry component, industry switching costs reduce migration beyond regional frictions, further attenuating estimates.

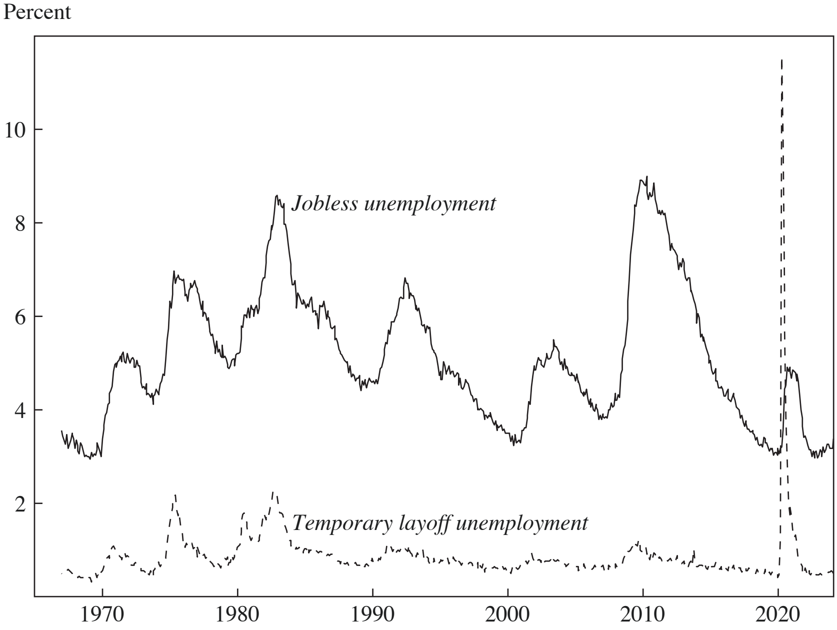
Fieldhouse, Munro, Koch, and Howard allude to a possibility that even if workers are very responsive to local shocks, there might be little incentive to relocate when potential alternative locations face similar shocks. Further investigation into understanding migration responses to local shocks might be a promising area for future research.

THE UNEMPLOYMENT RECOVERY FROM THE PANDEMIC RECESSION The newly constructed state unemployment series enable Fieldhouse, Munro, Koch, and Howard to study the evolution of states' unemployment recoveries from 1947 to the present. They find a convergence across states in both the speed and degree to which unemployment recovers after recessions and document the emergence of a national business cycle, which is experienced more uniformly across the United States.

The recovery from the pandemic recession, however, appears as an outlier from this pattern—the dispersion of the recovery speed across states is increased. In this comment, I take a closer look at the pandemic recovery and argue that a proper comparison of the pandemic recession to the previous recessions requires distinguishing between temporary layoff unemployment and unemployment due to other reasons. Once the distinction is taken into account, the recovery from the pandemic recession is not as dissimilar from the previous recoveries.

In Hall and Kudlyak (2022a), we analyze the historical behavior of unemployment and find that it comprises occasional sharp upward movements in economic crises and, at other times, an inexorable downward glide at a low but approximately constant proportional rate of about 0.1 log points per year. The rate of decline is approximately similar across the ten recoveries prior to the pandemic. In Hall and Kudlyak (2022c), we ask what can be behind these inexorable recoveries of unemployment. Despite high variations in monetary and fiscal policy, productivity, and labor force growth, there was little variation in the rate of decline of unemployment. Our thesis is that a natural force causes jobless job seekers to match with available jobs and to lower unemployment. The process is slow because a typical crisis breaks worker-firm employment relationships—separations

Figure 6. Temporary Layoff Unemployment Rate and Jobless Unemployment Rate



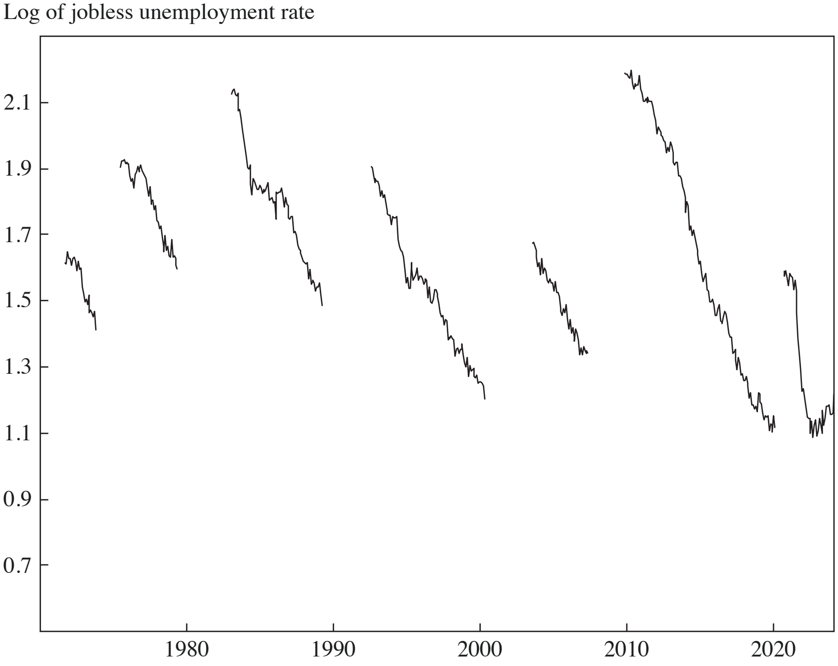
Source: Author's calculations using data from CPS, retrieved from Haver.
Note: Data through February 2024.

are permanent—and creating new, stable relationships is time-consuming and costly (Hall and Kudlyak 2019). That is, our thesis is that a self-recovery of the labor market after a large fraction of matches is destroyed by a recessionary shock is behind the inexorable recoveries of unemployment.

One distinct feature of the pandemic recession was a huge run-up in temporary layoff unemployment. In fact, the entire increase in unemployment during the March–April 2020 recession came from temporary layoff unemployment (Hall and Kudlyak 2022b). The unemployed on temporary layoff typically do not go through the search-and-matching process, so no firm-worker match capital is destroyed.

In Hall and Kudlyak (2022b), we show that to understand the labor market during the pandemic and its aftermath, one should examine separately temporary layoff unemployment and unemployment due to other reasons—jobless unemployment. Figure 6 shows temporary layoff unemployment as share of the labor force and jobless unemployment as share of the labor force. The two series sum up to the official unemployment rate.

Figure 7. Recoveries of Jobless Unemployment



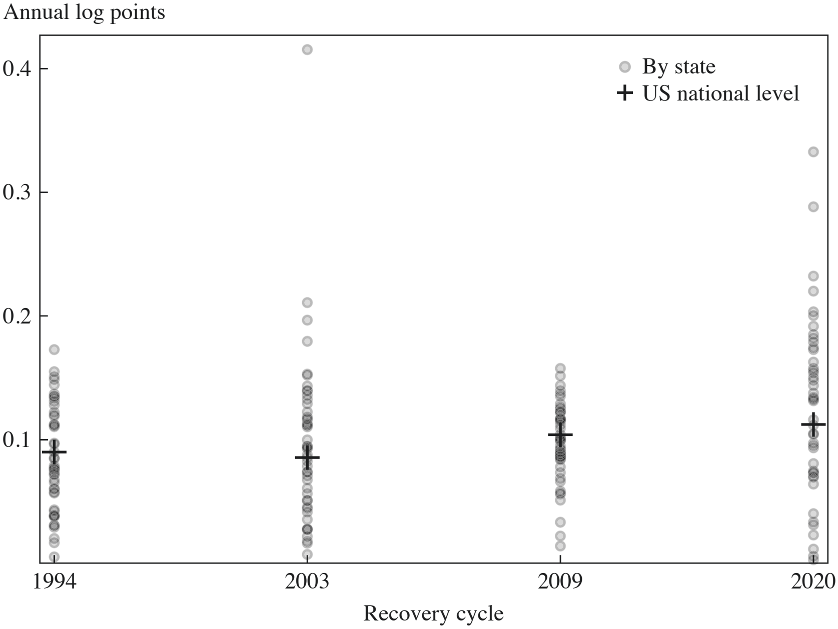
Source: Author's calculations using data from CPS, retrieved from Haver.
Note: Data through February 2024. The chronology of the recoveries is from Hall and Kudlyak (2022a).

Historically, a large fraction of unemployment was jobless (Wolcott and others 2020). For example, in the 2007–2009 recession, jobless unemployment reached 9 percent. The jobless unemployment rate increased slowly and peaked at 4.9 percent in November 2020.

Figure 7 shows recoveries of jobless unemployment. Specifically, it shows the logarithm of the jobless unemployment rate during the recoveries, defined from the month when the series reaches its peak to the month right before it starts increasing in a subsequent recession. The last data point in the figure is from February 2024. So far, the recovery of jobless unemployment from the pandemic recession is nonlinear, and it appears faster (e.g., the line is steeper) than the previous recoveries. Coincidentally, the average speed of the current recovery as of March 2024 is very similar to the 10 percent per year speed found for previous recoveries (figure 8).

One cannot read much into the average speed of the recovery, given the nonlinear nature of the recovery. However, to get some ballpark figures,

Figure 8. Recoveries of Jobless Unemployment, National and by State



Source: Author's calculations using data from CPS, retrieved from Haver.
Note: Data through February 2024. The chronology of the recoveries is from Hall and Kudlyak (2022a). The jobless recovery of 2020 starts in November 2020.

I calculate the dispersion of the average recovery speed of jobless unemployment across the US states (this is similar to figure 7 in the paper, which is constructed for total unemployment). These results are shown in figure 8. As can be seen, the dispersion of the average jobless unemployment recovery speed across states during the pandemic recovery is much smaller than the dispersion of the total unemployment recovery speed across states as shown in figure 7 in the paper.

CONCLUSIONS Fieldhouse, Munro, Koch, and Howard provide a significant contribution to the profession by creating state unemployment rate series for the postwar period. They show a consistent pace of state unemployment recoveries and a more uniform business cycle across states post-1960s. The recovery pace of jobless unemployment in the post-pandemic period aligns with historical patterns. Future research should focus on distinguishing between different types of unemployment to better understand recovery dynamics.

REFERENCES FOR THE KUDLYAK COMMENT

Ahn, Hie Joo, Bart Hobijn, and Ayşegül Şahin. 2023. "The Dual U.S. Labor Market Uncovered." Finance and Economics Discussion Series 2023-031. Washington: Board of Governors of the Federal Reserve System.

Auray, Stéphane, David L. Fuller, and Damba Lkhagvasuren. 2019. "Unemployment Insurance Take-Up Rates in an Equilibrium Search Model." *European Economic Review* 112: 1–31.

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity*, no. 1: 1–75.

Borusyak, Kirill, Rafael Dix-Carneiro, and Brian K. Kovak. 2023. "Understanding Migration Responses to Local Shocks." Working Paper. https://www.andrew.cmu.edu/user/bkovak/borusyak_dixcarneiro_kovak.pdf.

Gregory, Victoria, Guido Menzio, and David G. Wiczer. 2024. "The Alpha Beta Gamma of the Labor Market." Working Paper. https://www.stern.nyu.edu/sites/default/files/2024-04/GMW_Apr24.pdf.

Hall, Robert E., and Marianna Kudlyak. 2019. "Job-Finding and Job-Losing: A Comprehensive Model of Heterogeneous Individual Labor-Market Dynamics." Working Paper 25625. Cambridge, Mass.: National Bureau of Economic Research.

Hall, Robert E., and Marianna Kudlyak. 2022a. "The Inexorable Recoveries of Unemployment." *Journal of Monetary Economics* 131: 15–25.

Hall, Robert E., and Marianna Kudlyak. 2022b. "The Unemployed with Jobs and without Jobs." *Labour Economics* 79: 102244.

Hall, Robert E., and Marianna Kudlyak. 2022c. "Why Has the US Economy Recovered So Consistently from Every Recession in the Past 70 Years?" *NBER Macroeconomics Annual* 36: 1–55.

Wolcott, Erin, Mitchell G. Ochse, Marianna Kudlyak, and Noah A. Kouchekinia. 2020. "Temporary Layoffs and Unemployment in the Pandemic." Federal Reserve Bank of San Francisco Economic Letter, November 16. https://www.frbsf.org/wp-content/uploads/el2020-34.pdf.

GENERAL DISCUSSION Katharine Abraham pondered the issue that some of the movements in the claims-based data series the authors have constructed may reflect changes in state unemployment insurance program rules or administration rather than real changes in the labor market. While the authors have made an effort to evaluate the likely effects of some of these program changes, including changes in program coverage, the maximum duration of benefits, and the length of the waiting periods for receiving benefits, Abraham suggested that changes in other features of states' programs whose impact is harder to assess also may have affected the data. For example, the number of claims filed will reflect state rules about the earnings history required to establish eligibility for benefits. Eligibility also may depend on whether one is searching for full-time or part-time work or on other factors. Perhaps more important, the generosity of benefits and the overall administrative burden associated with obtaining benefits may affect whether people who in principle are eligible choose to apply. While it would be difficult to quantify the effects of changes in individual factors, Abraham noted that, since the Great Recession, the average ratio of insured unemployment to official unemployment has declined, with some states having been more affected than others.

On data issues, David Munro tried to evaluate these administrative issues and agreed with their importance. The authors have compared their data with the historical annual insured unemployment rate (IUR) data and have found that they do indeed correlate, which suggests that these administrative changes are not influencing the authors' findings across the business cycle.

Gerald Cohen questioned the authors' claim that there has been a decline in structural dispersion, specifically on a metropolitan level. Cohen has measured economic activity for the 150 largest metro areas in the United States and found an increased dispersion in outcomes over twenty years.[1] For the fifty largest metro areas, the dispersion of gross domestic product (GDP) has widened from 2002 to 2019, and it has continued to widen in the post-COVID-19 period. Cohen compares Nashville and Memphis: they have similar potential migration to the state but different outcomes. There is increased dispersion not just on a metro level but within cities on an industry level, such as in San Francisco in the past twenty years or so.

James Stock argued that increasing correlations over business cycles across states is not inconsistent with increasing trend dispersion—both can be true at the same time. Stock pointed out that the paper presents readers with a good opportunity to think about drivers and correlations of interstate migration.

Considering Cohen's comments, Robert Gordon contemplated the location of future unemployment caused by artificial intelligence (AI). Gordon posited that about a third of the labor force are those who create content looking at the computer screen, which mostly overlaps with the work-from-home group, while the remainder can be separated into two groups: those who produce goods (manufacturing, agriculture, mining, utilities) and contact services (retail and wholesale, transportation, most of the education sector, and most of the health care sector). Gordon hypothesized that AI

1. Kenan Institute of Private Enterprise, "The American Growth Project," https://kenaninstitute.unc.edu/american-growth-project/.

will not destroy most jobs in the latter two groups. He further commented that the content creating jobs are disproportionately located in the superstar cities, and he predicted AI will have a disproportionate impact on people with relatively high incomes. Gordon noted this would be a reversal from the China shock.

Munro explained that they are not speaking to the city level, although within-state and between-state differences are important. There are claims data available for metro areas, if researchers would like to pursue this within-state research question. On the point of industrial convergence, Andrew Fieldhouse referenced their paper, which found that most of the industrial convergence transpired between the 1940s through the 1960s. He noted that, in fact, there was a small reversal in the early 2000s, which is not necessarily inconsistent with the metropolitan statistical area (MSA) analysis. The authors focused on the transition away from manufacturing, particularly in the early postwar era.

Abraham suggested that the changing age distribution of the population may be at least partially responsible for the declining responsiveness of state-level migration to economic shocks. A worker who is early in their career is more likely to move in response to an economic shock; older workers and especially people who are over age 65 are less likely to respond by moving. The aging of the population associated with the Baby Boom thus could have affected the migration response.

Thinking about other data sets that go far back, and that could complement the data of the authors, Stock noted that he and Mark Watson looked at housing starts across states and found decreasing dispersion through increasing correlation in growth rates across states. Stock remarked that the decline in the manufacturing sector might be a structural reason for these patterns, noting there could be additional explanations in other markets.

Janice Eberly provided additional support, explaining that the share of home buyers more than fifty miles from their current home is at a record high. Eberly suggested this is consistent with prior comments on the increase in the average age of the population, meaning more retirees, and supply-side issues in the housing market.

Maurice Obstfeld asked whether the decline in labor mobility might be partially explained by the increase in political polarization in the United States. Obstfeld then walked through two possible narratives for the decline in labor mobility. One suggests economic structures are converging and states have increasingly similar business cycles, which may be indicative of political convergence down the road. The other possibility is that people are constrained in one way or the other and cannot move. Similar to previous comments, Obstfeld pondered whether we are becoming less mobile because people are older on average, or perhaps housing markets are affecting mobility. Thinking about who it may be that moves, Obstfeld suggested that to the extent that only those of the elite class are able to move, such dynamics could accentuate polarization. Obstfeld was puzzled on one note: researchers know employers have decreased temporary layoffs, so people should be more mobile because they are not tied to their original employer.

On migration generally, the authors agreed there are several factors in play. Munro pointed to recent research that shows housing markets and home prices matter in terms of migration patterns.[2] The authors did not know about the availability of housing start data at the state level, but they found it quite interesting that there is similar convergence across states over a similar timeframe.

More broadly, Gordon shared that he had been puzzled about the finding in the paper by Hall that unemployment declines uniformly across all the business cycles.[3] Specifically, the recovery in the wake of the financial crisis was very slow, whereas the opposite was true for the recovery in 1983 and 1984. Gordon explained that the answer to this puzzle is in realizing that, in the latter case, the initial recovery was very rapid but slowed subsequently. Averaging across multiple years, then, will yield a result showing a relatively slower recovery, akin to that following the financial crisis.

Robert Hall, in response to Gordon, pointed to the fact that the irregularity of recoveries as measured by GDP is much greater than for unemployment. There is a remarkable smoothness to unemployment as compared to other factors. Hall commented that even as driving forces differ, the labor market seems to produce this same result in unemployment during recoveries across business cycles, as explored in more detail in the paper by Marianna Kudlyak and him.[4]

Steven Davis agreed with Abraham's concern about the conceptual difference in the authors' measure of unemployment and that produced using the Current Population Survey (CPS). Davis explained that there used to

2. William W. Olney and Owen Thompson, "Determinants of Declining Internal Migration," working paper 32123 (Cambridge, Mass: National Bureau of Economic Research, 2024).

3. Robert E. Hall and Marianna Kudlyak, "The Inexorable Recoveries of Unemployment," *Journal of Monetary Economics* 131 (2022): 15–25.

4. Ibid.

be a data set on annual state-level unemployment rates dating back to the 1940s, but that these data have now been lost. Davis added that there was an active literature in the 1960s and 1970s discussing how best to estimate state-level unemployment rates using the CPS definition through synthesizing micro data from the CPS, state insured unemployment rates, and other sources. A different paper could rediscover or recreate these data and evaluate their quality.

SAMUEL G. HANSON
Harvard Business School

JEREMY C. STEIN
Harvard University

VICTORIA IVASHINA
Harvard Business School

ADI SUNDERAM
Harvard Business School

LAURA NICOLAE
Harvard University

DANIEL K. TARULLO
Harvard Law School

The Evolution of Banking in the 21st Century: Evidence and Regulatory Implications

ABSTRACT As revealed by the failures of three regional banks in the spring of 2023, bank runs are not a thing of the past. To inform the ongoing discussion of the appropriate regulatory response, we examine trends in the banking industry over the last twenty-five years. On the liability side of bank balance sheets, deposits—and especially uninsured deposits—have grown rapidly. On the asset side, there has been a notable shift away from the information-intensive lending traditionally associated with banks and toward longer-term securities such as mortgage-backed securities and long-term Treasuries. These trends appear to be related, in the sense that banks with the most rapid growth in deposits have seen the biggest declines in loans as a share of assets. Thus, while the banks that failed in early 2023 were arguably extreme cases, they reflect broader trends, especially among larger banks. We construct a simple model to help assess the main regulatory options to reduce the risk of destabilizing bank

Conflict of Interest Disclosure: Jeremy C. Stein and Daniel K. Tarullo are former members of the Federal Reserve Board of Governors. Adi Sunderam was a visiting scholar at the Federal Reserve Bank of Boston from January to June 2023, and Victoria Ivashina has been a visiting scholar there since 2015. Laura Nicolae is a PhD intern at the Federal Reserve Board of Governors, and her work on this paper was conducted outside of her internship work. The views presented in this paper are solely those of the authors and do not represent the views of the Federal Reserve Board or the Federal Reserve System. The authors did not receive financial support from any firm or person for this paper or, other than the aforementioned, 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 2024: 343–389

runs—expanding deposit insurance and strengthening liquidity regulation—and argue that the industry trends we document favor the latter option. Using the model, we offer some design considerations for modifying the liquidity coverage ratio so as to require banks to pre-position sufficient collateral—largely in the form of short-term government securities—at the Federal Reserve's discount window to ensure they have enough liquidity to withstand a run on their uninsured deposits. We also comment briefly on some other regulatory implications of our findings, including for interest rate risk regulation and merger policy.

The late winter and early spring of 2023 saw three of the four largest bank failures in US history, those of Silicon Valley Bank (SVB), Signature Bank, and First Republic Bank, on March 10, March 12, and May 1, respectively. This dramatic episode, and the failures in bank regulation that it revealed, naturally led to calls for a variety of regulatory changes. While we believe that this instinct toward reform is well motivated, in this paper we begin by taking several steps back. We try to sketch some of the broader forces that have been shaping the evolution of the banking industry, and of financial intermediation more generally, over the last quarter century. Our premise in doing so is that only by understanding how the economics of the banking industry have evolved can one begin to think sensibly about how regulation might be best adapted.

We organize our analysis around the two fundamental pillars of banks' business model: making information-intensive loans to borrowers who are risky and opaque and providing deposit-taking and transactions services. We then ask how developments in these two areas have affected banks in different size categories: (1) the largest global banks, the so-called global systemically important banks (G-SIBs), which currently have assets over \$700 billion; (2) regional banks, which, for the sake of concreteness, we classify as having assets between \$100 billion and \$700 billion today; and (3) smaller banks, which have assets less than \$100 billion today.[1]

1. Eight US bank holding companies currently qualify as G-SIBs: JPMorgan Chase, Bank of America, Citigroup, Wells Fargo, Goldman Sachs, Morgan Stanley, Bank of New York Mellon, and State Street. The first six of these institutions all have assets above \$1 trillion. The two custodian banks—Bank of New York Mellon and State Street—have assets below \$700 billion but are systemically important because of the central role they play in settling securities transactions. We recognize that there are no sharp dividing lines based on assets that can fully distinguish banks with different business models. So, for example, a number of banks with assets less than \$100 billion might have business models similar to those of some banks with assets over \$100 billion. See Financial Stability Board, "2023 List of Global Systemically Important Banks (G-SIBs)," press release, November 27, 2023, https://www.fsb.org/2023/11/2023-list-of-global-systemically-important-banks-g-sibs/.

The idea that banks—and financial intermediaries more generally—create value on the asset side of their balance sheets by screening and monitoring borrowers is perhaps the most venerable and widely accepted view in the academic literature. Diamond (1984) is the classic reference for this asset-side view of what makes banks special. However, the view that banks play a unique role in information-intensive lending has come under increasing pressure in recent decades, as nonbank institutions have steadily gained market share in lending to businesses. These nonbank players include securitization vehicles, mutual funds, and insurance companies that finance portions of syndicated loans—and, in more recent years, private credit funds and business development companies (BDCs) that lend to medium-sized firms. Moreover, it appears that the competition from private credit funds and BDCs has been felt most acutely by regional banks. By contrast, community banks, which tend to specialize in lending to much smaller firms, have been less affected by the growth of nonbank intermediaries.[2]

Another branch of the literature, beginning with Gorton and Pennacchi (1990), emphasizes the value that banks create on the liability side of their balance sheets, via their deposit-taking franchises. There are two logically distinct mechanisms at work here. The first is that some agents in the economy prefer holding absolutely safe assets as a store of value and that bank deposits are an especially good vehicle for providing this safety. Moreover, these same agents tend to be inattentive and will often accept below-market rates on their deposits, perhaps partially in exchange for the amenities provided by their bank—for example, friendly and accessible branch offices.[3]

A second source of value from deposits stems from their unique role in the payments system. In addition to being a safe store of value, bank deposits allow firms and households to transfer resources quickly and efficiently. A firm that uses its bank to handle transactions with its employees, suppliers, and customers is an example of this transactional function.

One of the most striking developments that we document over the last quarter century is a dramatic growth in the economy-wide ratio of bank deposits to GDP, with much of this growth coming from large uninsured deposits. Thus, very crudely put, the business of banking seems to be slowly moving away from a Diamond (1984) world and toward a Gorton and Pennacchi (1990) world. We reflect on some of the underlying causes

2. See Erel and Inozemtsev (2024) for an overview of the causes and consequences of the rise in nonbank lending.

3. To the extent that the value of a bank's deposit franchise comes from paying inattentive depositors less than the market rate (and adjusting for the cost of taking deposits), this is a private source of value but not a social benefit.

of this deposit growth in what follows, though to be clear, we do not have a single, encompassing explanation to offer. However, if one posits that the demand for payments services should scale roughly with GDP, the rapid growth in the ratio of deposits to GDP suggests that some of the action is coming from the safe-store-of-value motive, which might scale more naturally with wealth, rather than GDP.

Putting together these two trends—the migration of information-intensive business lending outside of the banking sector and the rapid growth of bank deposits—the inevitable consequence is a shifting of banks' asset portfolios toward categories where there is less of a presumption that they have a unique comparative advantage. Specifically, and this is especially true for the larger banks that have experienced the greatest competition from nonbank lenders, the share of securities in their portfolios has increased significantly in recent decades. These securities consist primarily of US Treasuries and agency mortgage-backed securities (MBS) whose payments are insured by the government-sponsored enterprises. These securities are free of credit risk, so the only risk that banks face in holding them is interest rate risk. In this sense, the larger banks are beginning to look more like long-term bond mutual funds than they did at the beginning of the century, albeit bond funds that have uninsured liabilities that can be withdrawn on demand at par rather than being equity financed. In what follows, we argue that this observation is of particular relevance when considering questions about whether and how regulators should modify deposit insurance coverage and bank liquidity regulation.

Of course, it can be artificial to frame things by simply contrasting theories wherein banks create value either on the asset side of their balance sheet or the liability side. There can be important synergies between the two sides of the balance sheet. For instance, in Diamond and Dybvig (1983) and Hanson and others (2015), banks can finance portfolios of illiquid loans more efficiently than other types of intermediaries so long as they can issue demand deposits that are not prone to destabilizing runs. With some liberties, this theory might be interpreted as warning that a failure to offer sufficiently broad deposit insurance coverage could interfere with the process of credit creation in the economy. This possibility highlights why it is critical to think about exactly what kinds of assets the marginal bank deposit is financing.

Alternatively, a synergy between the two sides of banks' balance sheets can arise if deposit taking, and the resulting need to hold a buffer stock of high-quality liquid assets as well as the associated access to the central bank's lender of last resort (LOLR) function, give banks a balance sheet–based edge over nonbank intermediaries in offering on-demand lines of credit (Kashyap, Rajan, and Stein 2002). Consistent with this view, we show that the one area of corporate lending where banks have not lost ground to nonbank intermediaries is in providing loan commitments to firms.

In what follows, we explore both time series and cross-sectional aspects of the abovementioned trends in banks' deposit-taking and lending behavior. We then turn to some of the policy implications of these trends. Here we begin by developing—with the aid of a simple model—a normative perspective on the design of bank liquidity regulation.

The bank failures of early 2023 highlighted a dramatic vulnerability with respect to liquidity risk, created by the combination of rapid growth of uninsured deposits and technological and social media innovations, which appear to have made bank runs more rapid and violent than ever before. As one extreme example, 94 percent of SVB's total deposits were uninsured on the eve of its failure, and 25 percent of its deposits were withdrawn in a single day, forcing its closure by regulators. Moreover, had it opened for business the next day, SVB told regulators it expected to see withdrawals of more than twice that amount in the following twenty-four hours (OIG 2023).

This episode lends urgency to the question of how such heightened run risk can best be mitigated. Two broad categories of options are: (1) increasing the scope of deposit insurance so most deposits are insured and hence unlikely to run; or (2) subjecting uninsured deposits to tougher liquidity requirements so the risk of runs poses a smaller threat to financial stability. Although both options are likely to deliver benefits in terms of mitigating run risk, they entail different costs. On the one hand, expanding deposit insurance would likely create additional moral hazard distortions and expose taxpayers to greater losses. On the other hand, tougher liquidity requirements—that is, requiring banks to hold a larger buffer stock of high-quality liquid assets to cover deposit withdrawals—might crowd out valuable information-intensive lending. The observation that, in both the time series and the cross section, the rapid growth in uninsured deposits has largely been used to fund growth in securities—and not in information-intensive lending—suggests that the costs of tougher liquidity requirements are lower, inclining us to this latter option.

Specifically, we propose a regulatory change that would require larger banks to back their uninsured deposits by pre-positioning collateral—largely in the form of short-term government securities—at the Federal Reserve's discount window. As we explain, the federal banking agencies could implement our proposed regulatory change by modifying current liquidity coverage ratio (LCR) requirements.

Of course, to the extent we have correctly identified some of the key underlying trends that are changing the business of banking, there may be reason to adjust other regulatory policies beyond just the pressing case of liquidity requirements. We focus briefly on two of these. One has to do with the treatment of interest rate risk in the regulatory capital regime. Currently, risk-based capital requirements do not account for the ex ante interest rate risk on long-duration securities like Treasury bonds and MBS. Moreover, even ex post, mark-to-market losses on these securities do not flow through to banks' regulatory capital, except for the largest G-SIBs. We argue that in a world where uninsured deposits make up a much larger share of banks' capital structure than in earlier decades, these policies need to be rethought.

Finally, we turn to merger policy. Our analysis suggests that the business model of regional banks may be particularly vulnerable to the broad forces that are likely to shape the banking industry in the coming years. Unlike the community banks, which focus on relationship lending to the smaller firms in the economy, regionals have lost a good chunk of their core business lending franchise to the nonbank sector. This leaves them disproportionately reliant on their deposit franchises for ongoing viability, at a time when the longer-run durability of these franchises also seems open to question. Moreover, regional banks may not have sufficient economies of scale and scope to compete with the handful of the very largest banks as technological innovation and artificial intelligence become more and more vital to profitability. Mergers within the midsize regional sector might be one helpful mechanism in moving the process of consolidation along, while minimizing harmful medium-term effects on competition and financial stability.

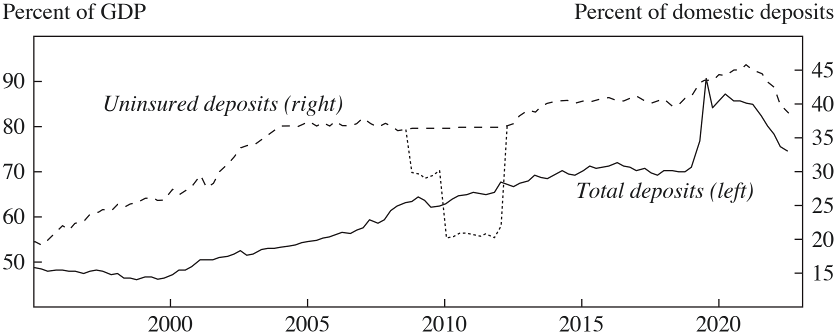
I. The Growth of Bank Deposits

Looking at quarterly data from 1995:Q4 to 2023:Q2, panel A of figure 1 plots the ratio of total deposits in US depository institutions to GDP alongside the ratio of uninsured domestic deposits to total domestic deposits at Federal Deposit Insurance Corporation (FDIC)-insured institutions.[4]

4. To facilitate consistency in this section, our total deposit series in figure 1 comes from the Financial Accounts of the United States and includes US-chartered depository institutions, US foreign banking offices, banks in US-affiliated areas, and credit unions. If we focus on US-chartered depository institutions—a universe that more closely matched the set of FDIC-insured institutions—the ratio of deposits to GDP rises from 41 percent in 1995:Q4 to 63 percent in 2023:Q2.

Figure 1. The Growth of Bank of Deposits

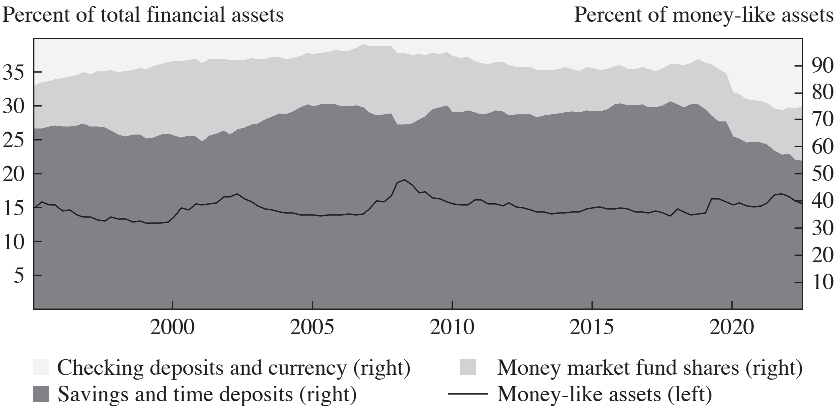
Panel A: Total deposits and uninsured deposits



Source: Total deposits are from the Financial Accounts of the United States and equal the sum of total checkable deposits and currency (FL793120005) and total time and savings deposits (FL703130005) minus the currency liabilities of the Monetary Authority (FL713120005). GDP is from FRED. The uninsured deposit share is from the FDIC's Quarterly Banking Profile.

Note: The solid line (left axis) shows total deposits at US depository institutions as a share of US GDP. This includes the deposit liabilities of US-chartered depository institutions, US foreign banking offices, banks in US-affiliated areas, and credit unions. The dashed line (right axis) shows the estimated fraction of domestic deposits that are uninsured at FDIC-covered institutions. For the uninsured share of deposits, we linearly interpolate the 2009:Q3 to 2012:Q4 values to remove the effect of the Transaction Account Guarantee (TAG) program, which lowered the uninsured share by temporarily expanding deposit insurance coverage.

Panel B: Households' holdings of money-like assets as percentage of their total financial assets



Source: Authors' calculations using data from table B.101 (Balance Sheet of Households and Nonprofit Organizations) from the Financial Accounts of the United States.

Note: The solid line (left axis) shows households' holdings of money-like assets—the sum of checking deposits, savings and time deposits, money market fund shares, and currency in circulation—as a fraction of the households' total financial assets. The right axis shows the fractions of these money-like assets that households hold in the form of checkable deposits and currency, time and savings deposits, and money market fund shares, respectively.

We linearly interpolate the uninsured share of deposits from 2009:Q4 to 2012:Q4 to visually smooth over the effect of the Transaction Account Guarantee (TAG) program, which temporarily lowered the uninsured share by providing unlimited insurance coverage on transaction deposits in the wake of the 2008 global financial crisis.[5] As panel A shows, deposits have grown rapidly relative to GDP over the past thirty years, with much of the growth coming from uninsured deposits. In 1995:Q4, deposits were 49 percent of GDP and the uninsured share was 20 percent. As of 2023:Q2, deposits are 75 percent of GDP and 39 percent of them are uninsured. Simply put, banks are much more deposit rich today than in past decades, but they are also far more exposed to the potential flightiness of uninsured deposits.[6]

This reliance on uninsured deposits is most pronounced for larger banks. As of 2023:Q2, 30 percent of domestic deposits in smaller banks—those with assets under $100 billion—are uninsured. For banks with assets over $100 billion but that are not G-SIBs, the corresponding figure is 39 percent. And for the G-SIBs, it is 51 percent. Indeed, across the latter two categories, 27 percent of banks have an uninsured deposit share than exceeds 50 percent.[7]

To shed some light on the forces driving these trends, figure A1 in the online appendix shows the evolution of a broader measure of money-like assets. Specifically, we decompose deposits into the sum of checkable deposits and savings and time deposits. To arrive at our broader measure of money-like assets, we then add the sum of currency in circulation and money market mutual fund shares. While there are cyclical fluctuations in this broader measure (e.g., money-like assets tend to rise relative to GDP during recessions and market downturns), money-like assets have trended steadily upward in recent decades, rising from 63 percent of GDP

5. The TAG program provided *unlimited* insurance on deposits held in noninterest-bearing transaction accounts for banks that chose to participate. The FDIC created this program in October 2008 using an emergency “systemic risk determination,” and it was in effect until the end of 2010. In mid-2010, Congress enacted a similar program for all banks that remained in effect until the end of 2012. See FDIC, “Temporary Liquidity Guarantee Program,” https://www.fdic.gov/banker-resource-center/temporary-liquidity-guarantee-program.

6. The uninsured share was also high from the advent of the FDIC in 1934 through the 1970s. However, this was arguably because, adjusted for inflation, insurance limits were much lower in those earlier decades. Thus, what is anomalous is today’s combination of a high uninsured share and a generous insurance limit in inflation-adjusted terms.

7. These figures are based on Call Reports data retrieved from Federal Financial Institutions Examination Council (FFIEC), “Central Data Repository’s Public Data Distribution,” https://cdr.ffiec.gov/public/PWS/DownloadBulkData.aspx. Using these data, we estimate that 41 percent of banks’ deposits were uninsured in 2023:Q2. The FDIC estimates that the uninsured share of domestic deposits was 39 percent in 2023:Q2 (figure 1, panel A).

in 1995 to 107 percent of GDP in 2023:Q2. Similarly, even though there are some noticeable cyclical shifts tied to the level of short-term interest rates, the shares of different money-like assets have been fairly stable.[8]

Next, using data from the Financial Accounts of the United States, figure A2 in the online appendix breaks down the holders of money-like assets.[9] Consistent with the well-documented rise in corporate cash holdings (Bates, Kahle, and Stulz 2009; Graham and Leary 2018), the cash holdings of nonfinancial firms and nonbank financial institutions have grown noticeably relative to households' cash holdings. Nonetheless, households still hold the lion's share of money-like assets, accounting for 61 percent as of 2023:Q2 as compared to 27 percent for nonfinancial and financial firms.

Notably, the quantity of deposits and other money-like assets, as well as the uninsured share of deposits, rose sharply following the onset of COVID-19 in 2020. Moreover, checking deposits have grown at record rates since 2020, while the growth in savings and time deposits has languished by comparison. Arguably, some of these recent shifts reflect the heightened precautionary motives associated with the pandemic and the fact that interest rates were at the zero lower bound. In addition, there is also clear evidence from account-level data at JPMorgan Chase that these abnormally large deposit balances are partially due to the outsize fiscal transfers to households during the pandemic (Wheat and Deadman 2023). Finally, Acharya and Rajan (2023) and Acharya and others (2024) have argued that the Federal Reserve's quantitative easing (QE) policies have led to an expansion of uninsured deposit financing, as banks have had to turn to uninsured deposits to fund their much-increased holdings of reserves. Collectively, these factors arguably explain these notable pandemic-era shifts, all of which have begun to reverse in recent quarters. But figure 1, panel A, makes clear that the upward trend in the deposits-to-GDP ratio as well as the uninsured share has been ongoing for decades, predating both the arrival at the zero lower bound and the initiation of QE policies in 2008 as well as the onset of the pandemic in 2020.

8. When the Fed raises its short-term policy rate, the rates that banks pay on checking and savings deposits lag well behind other money market rates (which generally move in lockstep with the Fed's policy rate). Thus, when the Fed raises rates, savers tend to gradually substitute away from lower-yielding checking and savings accounts and toward higher-yielding time deposits and money market fund shares. Conversely, when the policy rate is low, lower-yielding checking and savings deposits tend to grow more rapidly than time deposits. See Drechsler, Savov, and Schnabl (2017).

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

With respect to the factors that underlie these longer-term trends, we do not have any clear-cut evidence to offer. As noted above, the growth in the ratio of deposits to GDP could reflect a safe-store-of-value motive, which might scale more naturally with wealth, rather than with GDP. Consistent with this view, while total deposits have been growing as a fraction of GDP, panel B of figure 1 shows that households' holdings of deposits and other money-like assets have been quite stable relative to their total financial wealth, suggesting that households' portfolio allocation to money-like assets has been stable over time. Thus, the secular rise in the ratio of deposits to GDP is clearly linked to the secular growth in financial wealth relative to GDP.[10] That said, investors' willingness to hold their safe assets in the form of bank deposits paying less than a market rate—rather than in money market fund accounts, for example—might have been greater, all else being equal, due to the low level of interest rates the United States has experienced in recent decades.

Turning to the upward trend in the uninsured share of deposits, it stands to reason that the secular rise in household wealth inequality and the growth in corporate cash holdings both play some role in driving this trend. However, in the absence of account-level data, it is difficult to say whether or not these are important contributing factors.

II. The Rise of Nonbank Corporate Lending

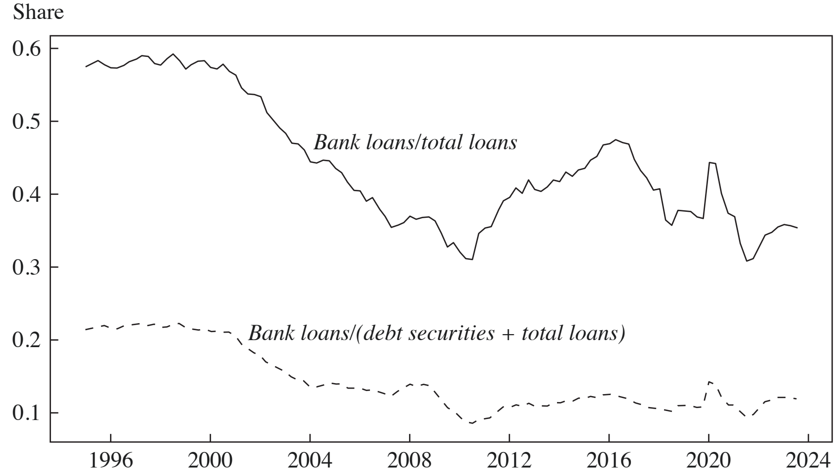
Figure 2 presents perspectives on the evolution of bank lending to nonfinancial businesses. Using data from table L.103 of the Financial Accounts of the United States, panel A focuses on nonfinancial corporate businesses. The solid line in the figure shows the ratio of bank loans to nonfinancial corporate businesses divided by total loans to these firms.[11] Importantly, the nonbank component of loans in the Financial Accounts data includes syndicated loans that are held by nonbank investors such as collateralized loan obligations (CLOs), mutual funds, insurance companies, and pension funds, but it *does not include* lending originated by private credit funds

10. Both Lopez-Salido and Vissing-Jorgensen (2023) and Buchak and others (2024) have previously noted that bank deposits have accounted for a stable fraction of household wealth in recent decades.

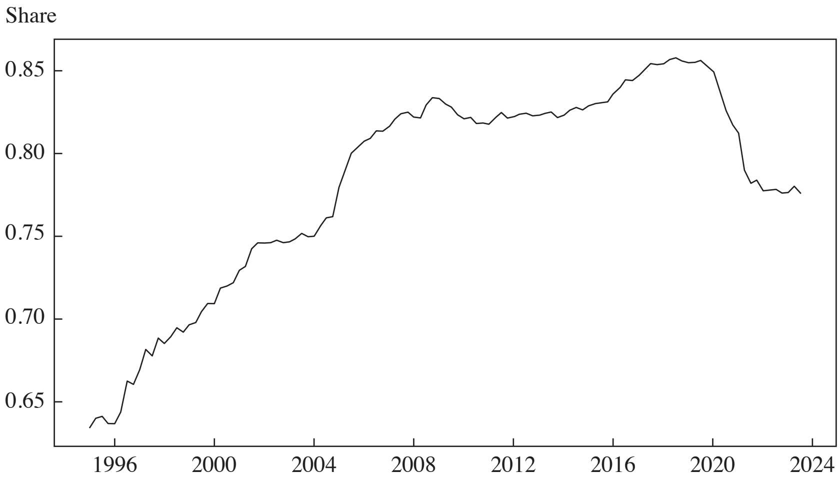
11. We exclude commercial mortgages from both the numerator and denominator since the Financial Accounts data do not break down commercial mortgages to nonfinancial corporations into those held by banks versus nonbanks.

Figure 2. Bank Lending to Nonfinancial Businesses

Panel A: Bank loans as shares of nonfinancial corporate liabilities



Panel B: Bank loans as share of loans to nonfinancial noncorporate businesses



Source: These figures are compiled using data from the Federal Reserve Board's Financial Accounts of the United States. Panel A uses series FL103168005.Q (bank loans), FL104123005.Q (loans), FL103165005.Q (mortgages), and FL104122005.Q (debt securities). Panel B uses series FL113168005.Q (bank loans) and FL113169005.Q (other loans and advances).

Note: In panel A, using data from table L.103 (Nonfinancial Corporate Business) from the Financial Accounts of the United States, the solid line shows bank loans (excluding mortgages) as a fraction of total loans (excluding mortgages) to nonfinancial corporate businesses. The dashed line adds corporate bonds to the denominator, plotting bank loans as a fraction of total loans plus corporate bonds. Panel B, using data from table L.104 (Nonfinancial Noncorporate Business) from the Financial Accounts of the United States, the solid line shows bank loans as a fraction of total nonmortgage loans to nonfinancial noncorporate businesses.

and business development companies (BDCs).[12] The dashed line in the figure adds corporate bonds and other debt securities to the denominator, showing bank loans to nonfinancial corporations as a share of all forms of credit (again, excluding loans from private credit funds and BDCs as well as mortgages).

Even before accounting for private credit funds and BDCs, panel A of figure 2 shows that banks currently provide a much smaller share of credit to nonfinancial corporations than they did at the turn of the century. As of 2023:Q3, bank loans account for only 35 percent of total nonmortgage loans and just 13 percent of total nonmortgage credit to nonfinancial corporations, down from 57 percent and 23 percent, respectively, in 2000:Q4. Naturally, banks also account for a similarly small fraction of the total growth in corporate credit over the past decade. From 2013:Q4 to 2023:Q3, bank loans to nonfinancial corporations grew by roughly \$700 billion. By contrast, nonbank loans to nonfinancial corporations grew by \$1.6 trillion and debt securities grew by almost \$3.1 trillion. Thus, bank loans account for 30 percent of the growth in total corporate loans and 13 percent of the growth in all corporate credit over the last decade.

Importantly, the trends seen for nonfinancial corporations do not show up when we look at lending to the noncorporate nonfinancial sector. As shown in panel B of figure 2, this sector, which can be thought of as capturing the smaller, unincorporated businesses in the economy, continues to be highly bank-dependent. The solid line in panel B displays the same construct as the solid line in panel A—bank loans to total nonmortgage loans—but for the unincorporated firms the bank share actually rises in the early part of the sample and has fluctuated between roughly 80 percent and 85 percent over the last twenty years. This divergence suggests that nonbanks are thus far not making meaningful inroads in lending to the smallest firms. This in turn implies that they pose less of a competitive threat to the small banks, whose lending business is largely dependent on relationships with these small firms. Rather, it is the lending model of the larger regional banks that appears to be most exposed to competition from nonbanks.

12. "Private credit" refers to nontraded commercial credit instruments that are originated and funded by nonbank institutions. Historically, private credit was used to finance midsize firms with revenues between \$10 million and \$1 billion. However, in recent years, private credit has been competing more directly with the syndicated loan market, which caters to larger firms. The biggest recent providers of private debt have been private credit funds and BDCs. Private credit funds are finite-horizon, closed-end funds that primarily invest in private credit instruments. BDCs also invest in private credit but are perpetual, closed-end funds that are financed using public equity and bond issues.

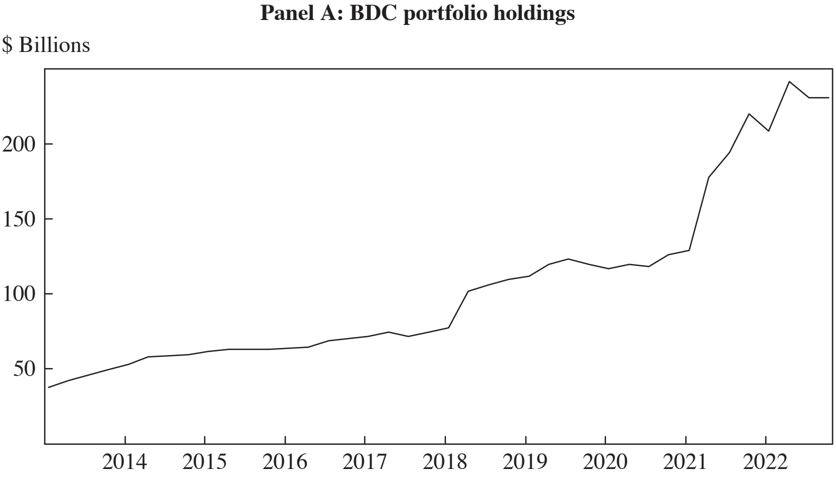
Returning to the corporate nonfinancial firms, the erosion in the bank share of credit to these firms—apparent in the solid line in panel A of figure 2—comes between 2000 and the onset of the global financial crisis in 2008. This is in part due to the rapid growth of nonbank leveraged lending during this period. Figure A3 in the online appendix documents the growth of the leveraged lending market by lender type over the 1996 to 2023 period. The leveraged lending market has always been dominated by nonbank financial institutions, including CLOs, mutual funds, insurance companies, and pension funds. Thus, rapidly growing leveraged lending represents an aggregate substitution away from bank-provided finance. Specifically, participation in the leveraged loan market by nonbank institutions grew from almost nothing in 2000, to about \$400 billion on the eve of the global financial crisis and stands at around \$1.2 trillion today.

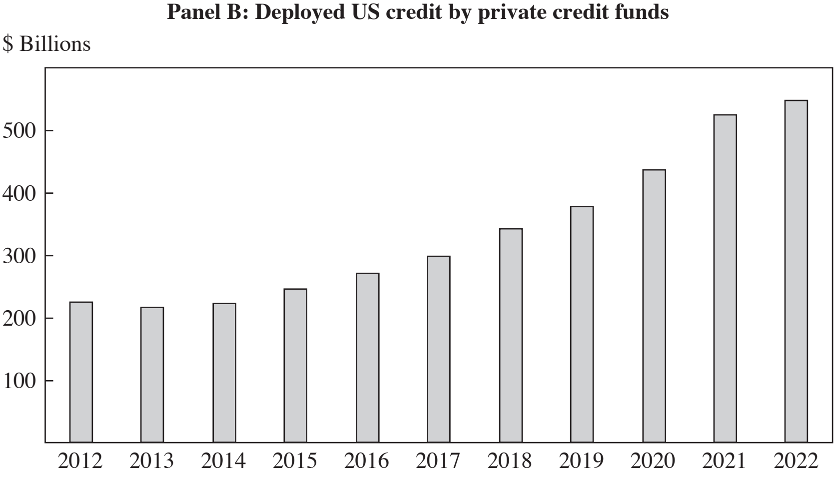
As noted above, panel A of figure 2 presents an incomplete picture of nonbank competition in lending to the corporate sector, because the data underlying the figure do not include private credit funds and BDCs, which grew very rapidly in the post–global financial crisis period. This can be seen in figure 3. Panel A of figure 3 plots loans held by BDCs. Total lending by BDCs has grown from about \$40 billion in 2013 to \$230 billion today. To benchmark these magnitudes, over the same period, total bank loans to nonfinancial corporations have grown by \$700 billion. So, the incremental market share captured by BDCs alone is economically quite significant.

Panel B of figure 3 plots an estimate of the deployed capital of US private credit funds. Since 2013, deployed capital by private funds—a concept broadly analogous to loans on their books—has grown by about \$300 billion. Thus, the combined lending to nonfinancial corporations from BDCs and private credit funds has grown by almost \$500 billion since 2013. This figure is roughly in the same ballpark as the \$700 billion increase in bank loans to nonfinancial corporations over the past decade. So, even excluding all other more established forms of nonbank finance to firms, such as the leveraged loan market and the corporate bond market, these two relatively new sources of nonbank credit alone are now very significant competitors in an important segment of the corporate lending market.

One place where banks have not lost any appreciable ground is when it comes to providing commitment-based revolving loans to corporations. According to Shared National Credit Program data as of 2022:Q2, banks hold over 97 percent of the \$1.4 trillion of outstanding syndicated revolving loans (OCC, Federal Reserve Board, and FDIC 2023a). By contrast, banks hold only 26 percent of the \$1.5 trillion of outstanding term loans. This implies that almost all the gains in market share that nonbank lenders

Figure 3. The Growth of Nonbank Lending





Source: Panel A is compiled using data from Pitchbook/LCD; panel B is compiled using data from Pitchbook.

Note: Panel A plots total loans held by BDCs. In panel B, US values of private debt assets under management (AUM) are estimated using Pitchbook data on global private debt AUM and applying a rolling five-year average of the US share of global fundraising.

have made in corporate lending have come in the market for installment credit. These findings are consistent with the view in Kashyap, Rajan, and Stein (2002) that deposit taking, and the resulting need to hold a buffer stock of high-quality liquid assets as well as the associated access to the central bank's lender of last resort function, gives banks a particular comparative advantage over nonbanks in supplying on-demand lines of credit.[13]

What explains these trends? At a high level there are two main forces that might explain banks' declining share of credit intermediation. First, the migration away from banks might be driven by advances in informational, contracting, and organizational technologies—for example, the development of securitization or new underwriting techniques by nonbanks. Second, the migration away from banks might be due to changes in financial regulation. Using a structural approach, Buchak and others (2024) find that changes in technology and the deepening of securities markets account for the considerable migration of credit intermediation away from banks that was witnessed from the 1970s to the 1990s. While this migration has continued since 2000—in part due to the heightened regulation of banks since the 2008 global financial crisis—they show that the rate of migration has decelerated. Reviewing the recent literature, Erel and Inozemtsev (2024) survey the evidence that heightened bank regulation has contributed to migration since 2008. At the same time, there is also strong evidence that nonbank lenders have been far more innovative and that these technological shifts have also contributed to migration since 2008 (Lerner and others 2024; Schneider, Strahan, and Yang 2023).

III. Implications for Bank Portfolio Shares

The combination of these two broad trends—rapid deposit growth and strong competition from nonbank providers of corporate credit—has, not surprisingly, left a mark on the composition of bank balance sheets. This is shown in table 1, which documents changes in banks' asset mix from 2000 to 2023. There are three panels in the table. Panel A examines the aggregate

13. The idea is that banks have a balance sheet–driven—as opposed to informational—advantage in extending revolving lines of credit. Since revolving loans can be drawn down on demand by borrowers, they have a similar contingent liquidity profile to demand deposits. Thus, to the extent that loan commitment drawdowns are imperfectly correlated with deposit withdrawals, a financial institution that combines deposit taking with commitment-based lending can economize on its costly buffer stocks of high-quality liquid assets. Empirically, loan commitment drawdowns tend to be strongly negatively correlated with deposit withdrawals in the time series, implying that banks have a significant advantage in making commitment-based loans (Gatev and Strahan 2006; Li, Strahan, and Zhang 2020).

Table 1. Bank Balance Sheet Shares (Percentage of Total Assets)

	(1) Loans (percent)	*(2) C&I loans (percent)*	*(3) Cash & securities (percent)*	*(4) Cash & securities < 3yrs (percent)*	*(5) Securities > 3yrs (percent)*	*(6) Reserves (percent)*
Panel A: All banks						
2000	60	17	27	n/a	n/a	0
2005	57	11	28	15	13	0
2010	53	9	30	18	12	5
2015	51	12	36	22	14	10
2020	52	13	35	21	14	7
2023	50	11	38	22	17	8
Panel B: Banks with less than $100 billion of assets (in $ 2023)						
2000	62	14	32	16	15	1
2005	64	11	29	15	14	0
2010	64	11	28	16	12	6
2015	65	12	29	14	15	5
2020	68	13	24	13	12	5
2023	65	12	28	12	15	5
Panel C: Banks with more than $100 billion of assets (in $ 2023)						
2000	61	20	24	n/a	n/a	0
2005	55	11	27	14	12	0
2010	49	9	33	21	12	5
2015	51	13	35	21	14	10
2020	51	14	36	21	15	8
2023	49	12	39	22	16	9

Source: Authors' compilation using data from FFIEC.

Note: The set of banks with over $100 billion of assets in 2023 dollars is not constant over time; it has grown as many banks have grown faster than inflation.

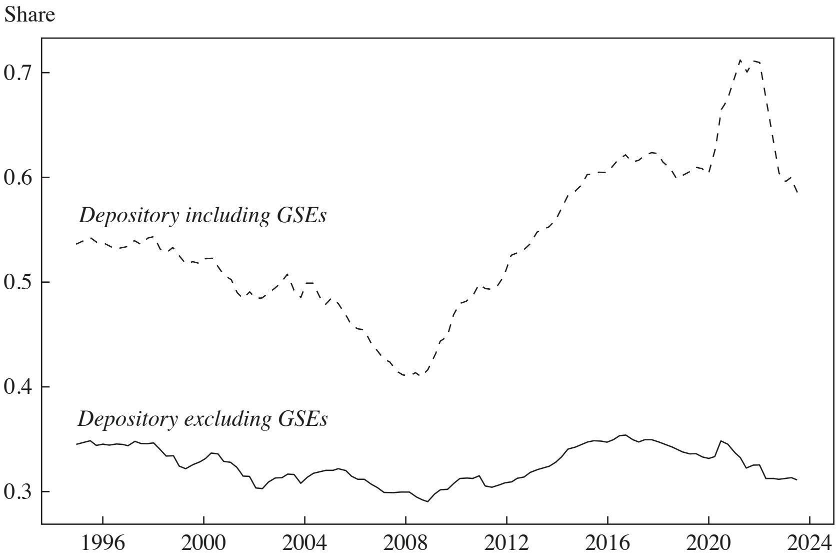
balance sheet of the entire banking sector over time and displays the share of bank assets represented by the following categories: (1) total loans; (2) commercial and industrial (C&I) loans; (3) total cash and securities; (4) cash and securities with a maturity of less than three years; (5) cash and securities with a maturity of greater than three years; and (6) central bank reserves. Panel B repeats the exercise but focuses only on those smaller banks with assets of less than $100 billion in 2023 dollars in each period. Panel C covers the complementary set, those larger banks with assets of greater than $100 billion.

Focusing on the panel C, we see that for larger banks total loans have fallen from 61 percent of assets in 2000 to 49 percent of assets in 2023. Moreover, almost all of this 12 percentage point decline is accounted for by the C&I category, where loans have fallen by 8 percentage points of assets, from 20 percent to 12 percent. Interestingly, however, this share has been roughly flat in the post–global financial crisis era, despite the very strong growth of private credit funds and BDCs, which one might have expected would have driven the bank portfolio share in C&I lending even lower. We suspect that the resolution to this apparent paradox is that overall loan demand, and hence aggregate lending volume, was very strong during this period of generally low interest rates and easy credit conditions. Mechanically, even if banks are losing a considerable share of the market for corporate loans, but at the same time the total size of the market is growing briskly, banks' volume of corporate lending may be holding up better than it otherwise would. Of course, a corollary of this reasoning is that if the growth of aggregate loan demand slows in the current higher interest rate environment and the nonbank providers of credit retain their higher market shares, banks' portfolio shares in C&I lending may decline even further.

The flip side of a reduced share of loans on bank balance sheets is an increased share of cash and securities. For the larger banks in panel C of table 1, we see that cash and securities have gone from 24 percent of assets in 2000 to 39 percent of assets in 2023, representing a quite dramatic reconfiguration of their balance sheets.[14] Furthermore—and this observation will be crucial when we turn to policy implications—even as total securities holdings have gone up, and even as these securities holdings are now increasingly funded with uninsured rather than insured deposits, the share of assets accounted for by securities with maturities over three years has actually increased somewhat, from 12 percent in 2005 to 16 percent today. This is important because, as the Silicon Valley Bank (SVB)

14. This trend is also emphasized by Stulz, Taboada, and van Dijk (2024).

Figure 4. Bank Share of One-to-Four-Family Residential Mortgages, Excluding Fed's Holdings



Source: Figure compiled using data from Flow of Funds (Federal Reserve Board): table L.109 (Monetary Authority), series LM713061705.Q; table L.110 (Private Depository Institutions) series LM703061705.Q; and table L.218 (One-to-Four-Family Residential Mortgages) series FL763065105.Q, FL893065105.Q, FL753065103.Q, FL743065103.Q, and FL473065100.Q. GSE = government-sponsored enterprise.

episode has taught us, an especially combustible mix is the combination of: (1) interest-rate risk coming from long-maturity securities holdings; and (2) a large proportion of runnable uninsured deposits (Drechsler and others 2023). Even if one believes that sticky and effectively long-duration insured deposits are a sensible way to fund long-duration securities, the same cannot be said for more run-prone uninsured deposits.

What are the long-duration securities that have become increasingly important on larger banks' balance sheets? Mortgage-backed securities (MBS) play a leading role. And indeed, the growth of their MBS holdings has helped turn banks into the leading private players in the mortgage market. This is illustrated in figure 4, which plots banks' share of the one-to-four-family residential mortgage market, where the total size of the market is defined excluding the Federal Reserve's holdings via its QE programs. There are two lines in the figure. The lower solid line captures banks' share of the whole loan mortgage market. As can be seen, banks are less prominent in terms of holding whole loans, with a market share that has fluctuated between roughly 30 percent and 35 percent over the last few decades but that shows no discernible trend.

The story looks very different when we examine the upper dashed line, which presents banks' share of the combined whole loan and agency MBS mortgage markets. Here the bank share soars from about 40 percent in 2008 to over 70 percent in 2021, before retracting somewhat to around 60 percent in 2023. In other words, their growth in MBS holdings is entirely responsible for banks' much increased presence in the overall mortgage market in recent years.

A first reaction to figure 4 might be that the rise in banks' share of the MBS market since 2008 is a mechanical reflection of the Fed's large purchases of MBS. This is not quite right. It is true that the Fed has taken a lot of MBS out of private hands, so that the bank share of the private market would mechanically grow even if bank holdings were not increasing in absolute dollar terms. But this fact still leaves the question why it is other nonbank private holders of MBS, such as bond mutual funds, that have been most willing to cede their MBS to the Fed. Said differently, bank demand for MBS has increased very strongly *relative* to MBS demand from other private investors over the last fifteen or so years. And these other investors are quite capable of intermediating agency MBS. Apparently, the combination of banks' eroding position in the corporate credit market and their large deposit inflows has given them a powerful appetite for MBS.[15]

Going back to table 1, it is instructive to compare the trends in balance sheet composition for the larger banks in panel C to those for the smaller (less than $100 billion in assets) banks in panel B. In sharp contrast to the larger banks, the smaller banks have not seen any noticeable decline in the share of either total loans or C&I loans on their balance sheets. For example,

15. Banks are overweight with MBS relative to a passive US government bond fund that owns Treasuries and agency-backed MBS in proportion to their outstanding market values. Specifically, Treasury and agency securities currently make up roughly 78 percent of banks' securities portfolio. Within this government securities bucket, banks currently hold 70 percent of their assets in agency MBS and the rest in Treasuries. By contrast, a value-weighted government bond fund would hold roughly 32 percent of its assets in agency MBS. (The numbers on bank portfolios are calculated from the Call Reports, and the bond fund figure is the ratio of outstanding agency MBS to outstanding marketable Treasury debt.) Although we cannot offer definitive proof, we suspect that banks' preference for MBS reflects the facts that MBS receive nearly as favorable regulatory treatment but offer higher yields than Treasuries. The analogy is not exact, but agency MBS are similar to callable Treasury bonds and thus offer a meaningful yield spread over Treasuries because MBS holders are short a valuable call option. However, since banks are typically concerned with the reported interest income on their securities—that is, banks care about yield and not simply total returns—they may perceive MBS as being more attractive than Treasuries; see Hanson and Stein (2015).

total loans are 62 percent of small bank assets in 2000 and 65 percent of small bank assets in 2023. Correspondingly, cash and securities are also roughly stable for small banks over the same period, going from 32 percent of assets to 28 percent of assets. This fits closely with the conclusion that we drew from the comparison of lending to nonfinancial corporate firms versus nonfinancial noncorporate firms in panels A and B of figure 2. Given that nonbank lenders have not gained significant market share in lending to the smallest firms in the economy, their growth has not made a discernible impact on the balance sheets of small banks. Instead, it is the larger regional banks whose business has been most disrupted by the increasing importance of nonbank credit providers.

IV. Cross-Sectional Evidence

A simple way to summarize our interpretation of the aggregate time series trends above is to say that, for the banking sector as a whole, deposit growth has outstripped growth in traditional lending opportunities in recent years. This contrasts with a situation where lending opportunities are growing rapidly, and banks must bid aggressively to raise additional deposits to finance an expansion of their lending portfolios. To further bolster our preferred interpretation, it is helpful to look in more detail at the cross section of banks. In table 2, we run the following cross-sectional regression over the 2010–2023 period:

$$\left(\frac{Y}{A}\right)_{i,2023} - \left(\frac{Y}{A}\right)_{i,2010} = \alpha + \beta \cdot \ln\left(\frac{Deposits_{i,2023}}{Deposits_{i,2010}}\right) + \varepsilon_i,$$

where A is total bank assets, and Y refers to a variety of specific asset categories (e.g., total loans, commercial and industrial [C&I] loans, cash, securities, etc.). Thus, we are asking how deposit growth over the 2010 to 2023 period has been correlated with changes in asset composition in the cross section of banks. Panel A of table 2 displays the results for the set of banks with assets over \$1 billion in current dollars, and panel B focuses on the smaller set of large banks (twenty-three observations) that currently have over \$100 billion in assets.

Looking first at panel B, we see that among the larger banks, more rapid deposit growth is correlated with a decline in the share of total loans to assets, the share of C&I loans to assets, and the share of C&I plus owner-occupied commercial real estate (CRE) loans to assets. Correspondingly, more rapid deposit growth is associated with a sizable increase in the share

Table 2. Regression of Change in Bank Asset Shares on Deposit Growth from 2010 to 2023

	(1) *Total loans*	*(2)* *C&I loans*	*(3)* *C&I & CRE loans*	*(4)* *Cash & securities*
Panel A: All banks				
Change in log deposits	3.572 (1.243)***	0.827 (0.575)	–0.808 (1.002)	–1.757 (1.079)
Observations	814	814	814	814
R^2	0.030	0.006	0.002	0.007
Panel B: Large banks				
Change in log deposits	–8.386 (1.374)***	–4.479 (0.887)***	–0.576 (1.180)	6.709 (2.267)***
Observations	23	23	23	23
R^2	0.259	0.229	0.003	0.194

Source: Call Reports from FFIEC.

Note: “C&I & CRE loans” includes all C&I loans plus loans secured by owner-occupied nonfarm nonresidential properties. Estimated at the regulatory high-holder or standalone bank level. Excludes Goldman Sachs, Capital One, Morgan Stanley, State Street, American Express Bank, and Discover Bank. Standard errors are robust to heteroskedasticity.

of cash and securities to assets. In terms of economic magnitudes, the point estimates imply that a one standard deviation increase in deposit growth is associated with a 4.1 percentage point decline in the ratio of loans to assets, offset by a 3.3 percentage point rise in the ratio of cash and securities to assets. Given the purely descriptive nature of these regressions, we are hesitant to read too much into the coefficient estimates. Nevertheless, they fit qualitatively with the inference we have drawn from the time series, namely that, especially among the larger regional banks, deposit growth has led to a reduced share of loans on the balance sheet and an increased share of cash and securities.

Panel A covers all banks and, importantly, weighs them all equally, so that the results are driven primarily by the smaller banks. Here, the patterns are directionally reversed, and the statistical significance is spotty. Now a one standard deviation increase in deposit growth between 2010 and 2023 is associated with a rise in the ratio of loans to assets of 2.3 percentage points and a decline in the ratio of cash and securities to assets of 1.1 percentage points.

We next turn to the role of uninsured deposits more specifically. It could be the case that uninsured deposits are particularly important for funding lending on the margin—perhaps because banks turn to the uninsured wholesale deposit market when their lending opportunities are too expansive to be funded by their retail deposit bases. This turns out not to be the case.

Table 3. Cross-Sectional Regressions of 2023:Q2 Asset Shares versus Uninsured Deposit Intensity

	(1) Loans	(2) Cash & securities	(3) Cash & securities ≤ 3yrs	(4) Cash & securities > 3yrs	(5) Reserves
Uninsured deposits/ assets	–0.133***	0.131***	0.072*	0.059*	0.052**
	(0.051)	(0.050)	(0.043)	(0.031)	(0.021)
Deposits/ assets	0.415***	–0.090	–0.309***	0.220***	–0.267***
	(0.113)	(0.112)	(0.096)	(0.036)	(0.076)
Observations	814	814	814	814	814
R^2	0.073	0.018	0.071	0.053	0.141

Source: Call Reports from FFIEC.

Note: The sample includes all banks with assets greater than $1 billion in 2023 dollars. A one-unit increase in the independent variable or dependent variable represents a 1 percentage point increase in the variable as a share of assets. Includes banks with at least $1 billion in assets (in 2023 dollars). Excludes Goldman Sachs, Capital One, Morgan Stanley, State Street, American Express Bank, and Discover Bank. Estimated at the regulatory high-holder or standalone bank level. Standard errors are robust to heteroskedasticity.

To see why, table 3 examines the cross-sectional relationship between balance sheet shares and the composition of deposits. Specifically, for a single cross section in 2023:Q2, and for the sample of the 814 banks with assets over $1 billion, we regress bank asset shares on uninsured deposits as a share of assets, controlling for total deposits as a share of assets. In other words, we are asking how asset composition changes as insured deposits are swapped for uninsured deposits, holding fixed total deposits.

The first column of table 3 shows that while the loans-to-assets ratio is positively correlated with the ratio of total deposits to assets, it is negatively correlated with the ratio of uninsured deposits to assets. A one standard deviation increase in the uninsured deposits-to-assets ratio is associated with a 1.8 percentage point decline in the loans-to-assets ratio. The remaining columns of the table show that this decline in loans is mirrored by a rise in cash and securities, with this increase roughly equally divided between reserves, cash and securities with maturities of three years or less excluding reserves, and securities with maturities greater than three years.

These results are again broadly consistent with the aggregate time trends documented above. In the aggregate, uninsured deposits have grown rapidly even as loans have declined as a fraction of assets. Similarly, in the cross section, high uninsured deposits are associated with less lending, not more. To some extent, this could reflect privately optimal liquidity management

Table 4. Regression of Bank Market-to-Book on Deposit and Loan Characteristics

	(1) *All banks*	*(2)* *Small banks*
Average deposit rate (pp)	–0.343***	–0.320***
	(0.076)	(0.069)
Average loan rate (pp)	0.043	0.051*
	(0.027)	(0.027)
Log deposits	0.581***	0.655***
	(0.120)	(0.121)
Log loans	–0.121	–0.173
	(0.114)	(0.114)
Log branches	–0.096*	–0.115**
	(0.048)	(0.046)
Log noninterest expense	–0.394***	–0.404***
	(0.116)	(0.114)
Log employees	0.017	0.060
	(0.098)	(0.092)
Bank-year observations	3,304	3,137
Within R^2	0.077	0.090

Source: Call Reports (retrieved from FFIEC) and S&P Capital IQ.

Note: This table reports annual panel regressions of a bank's market-to-book ratio on its deposit and loan characteristics. The regressions include year fixed effects and are estimated over the 2010 to 2023 period. All regressions exclude banks with less than $1 billion in assets in 2023 dollars as well as Goldman Sachs, Capital One, Morgan Stanley, BNY Mellon, State Street, American Express Bank, and Discover Bank. All banks owned by the same bank holding company in a particular year are collapsed into a single observation. "Small banks" include banks with $1 to $100 billion of assets in 2023 dollars. Standard errors are clustered by quarter and bank holding company (i.e., regulatory high holder).

in a world with deposit-led growth and modest lending opportunities. Banks flush with more uninsured deposits might be mindful that these deposits are potentially flighty and therefore hold larger liquidity buffers. However, it is worth noting that these liquidity buffers are held largely in the form of longer-maturity securities, and indeed, as uninsured deposits go up the cross section, so too does the share of longer-maturity securities on the balance sheet. As argued by Drechsler and others (2023), if uninsured deposits are vulnerable to run risk, this run risk may actually be exacerbated to the extent that these deposits are funding long-duration securities.

Finally, in table 4 we examine the importance of banks' deposit and lending franchises for bank equity valuations using a simplified version of the empirical strategy from Egan, Lewellen, and Sunderam (2022). The goal is to assess the degree to which each activity contributes to the private value of banks as seen by their shareholders. This private value may not be perfectly aligned with the social value banks create, but it is directly measurable from equity valuations. Table 4 considers a sample from 2010 to 2023 and estimates panel regressions of the form:

$$\left(\frac{M}{B}\right)_{it} = \alpha_t + \beta_D \bullet r_{it}^{Deposit} + \beta_L \bullet r_{it}^{Loan} + \gamma_D \bullet \ln\left(Deposits_{it}\right) + \gamma_L \bullet \ln\left(Loans_{it}\right) + \boldsymbol{\delta}'\mathbf{x}_{it} + \varepsilon_{it},$$

where $(M/B)_{it}$ is the market-to-book ratio of bank i in year t, $r_{it}^{Deposit}$ is the average net-of-fee rate that bank i pays its depositors, and r_{it}^{Loan} is the average rate that it earns on its loans. The regression asks how much a decrease in deposit rates or an increase in loan rates raises bank equity valuations, holding fixed the scale of deposit taking and lending. We include year fixed effects so that the coefficients are identified from cross-sectional variation across banks in a given year rather than variation over time.

Column 1 of table 4 examines all publicly listed banks with assets over \$1 billion in current dollars. The coefficient on deposit rates, β_D, is negative and significant, indicating that, as expected, banks that pay their depositors lower interest rates have higher equity valuations. The coefficient on loan rates, β_L, is positive: banks that earn higher rates on their loans also have higher valuations. However, the coefficient is close to zero in magnitude and is insignificant, suggesting that for all banks, the deposit franchise contributes far more to stock market value than the lending franchise.[16] Column 2 shows that we obtain similar results if we restrict attention to the subset of banks with assets of less than \$100 billion.

The difference between the value created by deposits and the value created by the lending business can be better understood by decomposing the market-to-book ratio into the price-to-earnings ratio and the earnings-to-book ratio (return on equity). In untabulated results, we find that lower deposit rates and higher loan rates both increase the earnings-to-book ratio—that is, both increase bank profits. However, lower deposit rates do not affect the price-to-earnings ratio, while higher loan rates are correlated with lower price-to-earnings ratio. In other words, stock market investors treat banks with higher loan rates as riskier and hence penalize their valuations accordingly. But they do not treat banks with lower deposit rates in the same way.

16. Using a more sophisticated empirical approach, Egan, Lewellen, and Sunderam (2022) reach a similar conclusion.

To summarize our empirical findings: over the last twenty-plus years, banks have seen rapid growth in their deposits, with much of this growth coming from uninsured deposits. At the same time, larger banks—those broadly categorized as regional banks—have faced increasing competition on the lending side from a variety of nonbank players, including most recently the fast-growing private credit and business development companies (BDC) sectors. As a result of these two forces, the asset portfolios of the regional banks have shifted significantly away from lending and toward holdings of long-term securities, specifically long-term Treasury bonds and MBS. These time series patterns also have analogs in the cross section, where we find that those banks with the fastest growth of deposits in recent years have seen the biggest declines in lending as a share of assets and the biggest increases in cash and securities as a share of assets.

In what follows, we ask how these observations about the evolution of the banking system should shape one's views toward bank regulation in general, and particularly toward liquidity regulation—that is, regulatory efforts to mitigate the run risks posed by much increased levels of uninsured deposits in the system.

V. Policy Implications

We now turn to policy implications. We discuss three topics: (1) the design of deposit insurance and liquidity regulation—specifically, how best to deal with the run risk created by large amounts of uninsured deposits in the banking system; (2) how capital regulation might be adjusted to deal with interest rate risk on banks' securities holdings; and (3) merger and competition policy. The first of these, deposit insurance and liquidity regulation, involves some subtle trade-offs, and we sketch a simple model to help clarify the issues.

V.A. Deposit Insurance and Liquidity Regulation

As noted above, 39 percent of all domestic deposits currently held in US banks are uninsured, an increase of 19 percentage points from 1995. And for banks with more than \$100 billion in assets, 27 percent of banks have uninsured deposits greater than 50 percent of their total domestic deposits. The bank failures of early 2023 highlighted the run risks associated with large amounts of uninsured deposits, and it now seems clear that technology and social media have, in certain circumstances, made these uninsured deposits more vulnerable to extraordinarily rapid and intense runs (Benmelech, Yang, and Zator 2023; Cookson and others 2023; Koont, Santos, and Zingales 2023).

The question we take up in this section is how best to address the run risk associated with this large volume of uninsured deposits. Our basic premise is that increased equity capital requirements alone, while helpful, are not sufficient for this task. There also needs to be a distinct and robust liquidity-oriented regime to complement capital regulation.

One obvious way to reduce the run risk associated with this high current level of uninsured deposits would be simply to expand the scope of deposit insurance coverage. As recently detailed by the Federal Deposit Insurance Corporation (FDIC), there are various options under this umbrella, from raising the deposit insurance limit somewhat from its current value of $250,000, to fully insuring business payment accounts, all the way to fully insuring all domestic deposits (FDIC 2023). Proponents of more aggressive versions of this approach sometimes argue that because uninsured depositors rarely are subject to losses in bank failures, these deposits are already de facto insured. So, the argument goes, one might as well make this insurance explicit and thereby eliminate run risk. Further, extending insurance to all deposits would entail banks paying higher deposit insurance premia, thereby forcing at least partial internalization by banks of the associated costs.

As a practical matter, it seems unlikely that Congress will expand deposit insurance coverage, at least in the foreseeable future. Thus, a response to run risk will almost surely need to be fashioned under the existing authority of the federal banking agencies. Political constraints aside, though, there are potentially important costs associated with a significant expansion of deposit insurance. Because deposit insurance can never be perfectly risk sensitive, expanding coverage will arguably create some additional moral hazard costs.[17] These costs could arise because deposit insurance distorts banks' ex ante risk-taking decisions in normal times—for example, by encouraging banks to invest in excessively risky assets—or banks' decisions after they have suffered large losses—for example, by allowing zombie banks to either lumber on or, even worse, to gamble for

17. Of course, the FDIC should strive to minimize the extent of moral hazard by making the insurance regime appropriately risk sensitive. However, since asset risk is not observable and since banks will arguably always know more about the risk of their assets than the FDIC, deposit insurance entails moral hazard costs. Thus, policymakers need to solve the second-best problem that involves trading off the run-stopping benefits of deposit insurance against its moral hazard costs in terms of distorting banks' decisions relative to the first best.

resurrection.[18] In its May 2023 review of options for deposit insurance reform, the FDIC also evinced concern about the impact of such a change in policy on the adequacy of the deposit insurance fund and the dynamics of wholesale funding markets (FDIC 2023).

An alternative approach to reducing run risk is to strengthen liquidity regulation by, for example, modifying the liquidity coverage ratio (LCR) to require uninsured deposits to be largely backed with Treasury bills (T-bills) and other short-term Treasuries. The LCR, which currently only applies to very large banks, requires that banks maintain sufficient high-quality liquid assets (HQLA) to cover their anticipated net cash outflows over a thirty-day period of stress.[19]

How one feels about this approach will naturally be colored by how one interprets the evidence we have presented above. At one extreme, if one believes that, at the margin, the banking system is raising uninsured deposits and largely investing them in long-term securities such as mortgage-backed securities (MBS), such an approach would seem relatively attractive. Having banks make investments in long-term securities is arguably a zero net present value (NPV) activity from a social perspective, since the intermediation of long-term securities can be efficiently carried out by bond mutual funds, without creating the severe run risks associated with uninsured deposit funding.[20] At the other extreme, if one has more of a Diamond and Dybvig (1983) view and believes that, even at the margin, wholesale bank

18. Even banks that are deeply insolvent often manage to stay above their regulatory capital minimums—and hence avoid intervention from forbearance-inclined regulators—given the backward-looking nature of accounting-based measure of equity capital. Although it would be a stretch to argue that uninsured depositors exert discipline on banks in the normal course of business, in many cases the event that forces an economically unviable bank to be shut down is a run by uninsured depositors. The savings and loan (S&L) crisis of the 1980s and early 1990s is a useful lesson in this regard, as many highly deposit-insured S&L institutions kept operating for many years in a zombie state, gambling for resurrection while increasing their losses and the ultimate costs to taxpayers.

19. The LCR specifies the eligible HQLA and, as discussed below, projects the anticipated net cash outflows during the thirty-day stress period based on an assumed run-off rate for each type of liability on the bank's balance sheet. At present, only banks with assets greater than $700 billion (or short-term funding greater than $75 billion) are subject to the full LCR, which requires them to hold enough HQLA to cover 100 percent of thirty-day stressed outflows. Depending on their levels of weighted short-term wholesale funding, banks with assets between $100 and $700 billion are subject to either a reduced LCR requirement or no LCR requirement at all.

20. Some have argued that, as in money market funds, investors in bond funds may enjoy a first-mover advantage in redeeming their shares during periods of stress. However, even those who agree with this view have not suggested the run risk is anything like that affecting a bank, which promises redemption at par on a first-come first-served basis.

deposits remain a uniquely efficient way to fund information-intensive credit provision, one is naturally going to be more sympathetic to expanding insurance coverage rather than leaning against the growth of wholesale deposits.

To clarify these issues and formulate a more specific proposal, we develop a simple model of a representative bank that initially funds itself in significant part with uninsured deposits and can invest in three assets: information-intensive loans (i.e., assets that are risky and illiquid), longer-term securities (i.e., assets that are risky but liquid), and short-term T-bills (i.e., assets that are both safe and liquid). The first goal of the model is to weigh the merits of expanded deposit insurance versus a modified LCR in dealing with deposits that are currently uninsured. An obvious proposition is that we should tilt in the direction of a modified LCR if expanding deposit insurance creates significant additional moral hazard or fiscal costs.

A somewhat more subtle proposition—one in the spirit of our empirics—is that an LCR rule is more costly when banks have a lot of positive-NPV lending opportunities, since forcing them to hold liquid assets to comply with the LCR will crowd out more valuable lending. If, as the data suggest, banks now have more uninsured deposits relative to their lending opportunities, an LCR rule looks more attractive compared to expanded deposit insurance.

Another goal of the model is to inform the design of the modified LCR. The model speaks to some of the key questions in adapting the LCR, including whether all uninsured deposits should be fully backed, what assets should qualify as backing for those deposits, and how liquidity regulation should interact with discount window lending.

To the extent that bank lending creates social value that is not equally available outside the banking system, the model suggests that the liquidity coverage requirement on uninsured deposits should be calibrated carefully so as to not overly constrict bank lending. At the same time, the model is quite clear in saying that it is problematic to back uninsured deposits with long-duration securities rather than T-bills and other short-term Treasuries. This is because we assume that long-duration securities can equally well be intermediated outside the banking system with less run risk by, for example, bond mutual funds. We use these implications, along with some considerations not addressed in the model, to put forward a framework for developing a more robust LCR.

MODEL ASSUMPTIONS The version of the model that we sketch here is deliberately kept very simple, with several shortcut assumptions made to minimize the required algebra and keep the focus on the policy implications.

We consider a representative bank—one of many identical banks—that operates at fixed scale and with a fixed capital structure: it has equity of E, small, insured retail deposits from households of D_R, and large wholesale deposits from firms of D_W. By fixing the capital structure in this way, we are implicitly assuming a frictional social cost of using additional equity financing. Otherwise, the problems that we address here could be solved at zero social cost simply by making the bank finance itself with a large quantity of equity. In that case, it could always lend at the first-best level while still holding enough liquid assets to buffer any amount of deposit outflows. So, while it is implicit, the constraint on equity is playing an important role.[21]

On the asset side, the bank can: (1) make loans of L; (2) hold longer-term risky securities in amount S; and (3) hold short-term, very low-risk securities—which we refer to as "T-bills" for simplicity—in amount B. So, the bank's initial balance sheet constraint is that $L + S + B = D_R + D_W + E$. There are three dates: At time zero, the bank chooses its asset mix. At time 1, there is an interim signal about the payoffs on the loans and the securities. With probability p, there is a bad signal. For loans, the bad signal implies that the expected time 2 payoff on the loans has declined to $F_L L < L$, and there is now a nonzero probability of an extremely bad crisis state in which the loans will only pay off some very small amount $0 \leq z_L L < F_L L$. We will begin by considering the limiting case where $z_L = 0$, but we will later ask how things change when $z_L > 0$. Similarly, for securities,

21. Why is bank equity costly? There are many reasons why it is *privately* costly for banks to rely on equity financing. However, many of these private costs do not qualify as social costs: while they affect the division of the economic pie between bank equity holders and other agents, they do not have an impact on the total size of the pie (Admati and others 2013). For example, the tax disadvantages of equity are a private but not a social cost. Of course, since deposit taking is socially valuable, equity capital requirements that limit banks' ability to accommodate the demand for deposits may be socially costly. However, this does not explain why it would be socially costly for banks to issue large amounts of equity to expand their holdings of high-quality liquid assets. In that case, banks could both lend and take deposits at the first-best level while holding enough liquid assets to meet deposit outflows. In this regard, one possible social cost of equity might arise from the agency problem between bank managers and outside investors, with the idea being that debt—particularly short-term debt—helps discipline managers, thereby increasing the size of the pie (Diamond and Rajan 2001). However, even if one believes that the direct social costs of bank equity are small, a substantial increase in bank equity capital requirements might still be costly for society. This is because, in attempting to economize on the private costs of equity, lending activity could flow out of banks and into other more lightly regulated areas, thereby posing threats to financial stability (Hanson, Kashyap, and Stein 2011).

the bad signal implies that the expected payoff has declined to $F_S S < S$, perhaps because interest rates have risen in the bad state. The key distinction between loans and securities is that while both can lose value at time 1, the securities are nonetheless perfectly liquid in that they can be sold for their full expected value at time 1. By contrast, as we explain in more detail shortly, the loans are illiquid at time 1, and selling them involves accepting a fire-sale discount relative to fundamental value. At time 2, all payoffs are realized.

We assume that there is a first-best level of loans L^{FB}. Any amount of lending L up to this level creates social surplus of πL, where $\pi > 0$ is a constant; beyond this point, lending creates no incremental social value. We further assume that $L^{FB} < D_R + D_W + E$, so that even if the bank is doing the first-best level of lending, it will hold some T-bills or securities. Thus, we are focusing on deposit-rich banks—that is, banks whose ability to lend is not constrained by the availability of deposits. This is consistent with the findings from our empirical work. At the same time, we assume that $L^{FB} > D_R + E$. This creates a meaningful tension, since if we require that wholesale deposits be fully backed with liquid T-bills or securities, this will push lending below the first-best level.

THREE SIMPLE POLICY OPTIONS By assumption, the retail deposits of D_R are always insured. We then begin our analysis by contrasting three simple policy options for dealing with the wholesale deposits. To be clear, these three options are effectively polar extremes and are intended to highlight the trade-offs at play in the starkest way.

Option 1: Full expansion of deposit insurance. In this case, the large wholesale deposits of D_W are fully insured. As a result, there are no runs at time 1 and no liquidity-based reason for the bank to forgo lending in order to hold an excess buffer stock of liquid assets. So lending is at the first-best level of L^{FB}, and the only social cost is that the increased deposit insurance leads to some additional moral hazard or fiscal cost, which imposes a social cost of $X > 0$. One interpretation of this cost, which is in the spirit of Diamond and Rajan (2001), is that because there is no run in the bad state at time 1, insolvent banks do not get shut down by regulators and become over-leveraged zombies who make bad lending decisions. So, the cost is only realized at time 1 in the bad state of the world and represents a form of excessive forbearance. Alternatively, a bank that is fully insured may make bad ex ante decisions, that is, take on negative-NPV risky bets at time zero.

Option 2: No expansion of deposit insurance, no liquidity regulation. In this case, the wholesale deposits remain uninsured, and the bank freely chooses its asset mix without any regulatory constraints. Suppose it picks quantities L^*, S^*, and B^* for loans, securities, and bills, respectively. Here one potential cost is that, because of the risk of insolvency, uninsured depositors necessarily run at time 1 upon observing the bad signal; this is their only way of assuring that they will be paid in full. And these depositors may have to be accommodated by fire-selling some illiquid loans, to the extent that the market value of the liquid securities and bills is not enough to cover all the uninsured deposit outflows. Although the loans of L^* have an expected value of $F_L L^*$, if they are fire-sold at time 1, they fetch only $k_L F_L L^* < F_L L^*$, where $k_L < 1$ is the fire-sale discount. To pay off all the wholesale depositors at time 1, the bank has to sell a fraction Δ_L of its loans such that $\Delta_L k_L F_L L^* + F_S S^* + B^* = D_W$. The private cost to the bank is the expected value of fire-sale losses on its loans: $p\Delta_L(1 - k_L)F_L L^* = p(1/k_L - 1)D_W - B^* - F_S S^*)$. Because the bank internalizes these fire-sale losses, it will seek to mitigate them by holding liquid assets and doing less lending. Thus, even without an LCR, the bank will choose to set $L^* < L^{FB}$. That is, the bank will self-impose some form of liquidity buffer policy.

To see what this self-imposed liquidity buffer looks like, suppose for the moment that the bank sets $S^* = 0$—that is, that the buffer is held entirely in T-bills as opposed to longer-term securities, so the bank's balance sheet constraint implies $(L - D_R - E) = (D_W - B)$. At an interior optimum, where the bank is indifferent between loans and bills, the marginal value of an additional loan must equal the fire sale—preventing benefit of an additional bill, which implies that $\pi = p(1/k_L - 1)$. We assume that $1/k_L$ is determined in equilibrium by the fire sales of all banks and is increasing in the quantity of fire sales $D_W - B$. Letting $h[D_W - B]$ denote the private costs of fire sales, where $h'[D_W - B] = p(1/\mathrm{k}_\mathrm{L} - 1) > 0$ and $h''[D_W - B] > 0$, the outcome in the unregulated case where the bank chooses the buffer satisfies $\pi = p(1/k_L^* - 1) = h'[D_W - B^*] = h'[L^* - D_R - E]$, where $k_L^* < 1$ is the equilibrium fire-sale discount and $L^* < L^{FB}$.

The need for a stricter regulatory LCR rule arises to the extent that fire sales of loans create social costs that are not internalized by individual banks. To capture these in a simple way, assume that when the bank liquidates $(D_W - B)$ loans to cover uninsured deposit withdrawals, the expected private costs are $h[D_W - B]$, but the expected social costs are $(1 + \phi)h[D_W - B]$, where $\phi > 0$. In other words, we assume that these fire sales impose some financial stability costs that the bank does not fully internalize (e.g., a negative effect on the balance sheets of other firms holding the affected assets or a negative effect on real investment). This creates a motive for a regulator to require the bank to hold more T-bills and engage in less lending than the bank would choose if left to its own devices. Specifically, the planner wants the bank to make loans L^{**}, where $\pi = (1 + \phi)p(1/k_L^{**} - 1) = (1 + \phi)h'[L^{**} - D_R - E]$, implying that $D_R + E < L^{**} < L^*$.

Thus, the total social cost of the unregulated market outcome is given by $\pi(L^{FB} - L^*) + (1 + \phi)h[L^* - D_R - E]$ and consists of both the cost in terms of forgone lending and the social fire-sale cost. By definition, this is greater than the social cost that the planner could achieve using optimal LCR regulation, which is $\pi(L^{FB} - L^{**}) + (1 + \phi)h[L^{**} - D_R - E]$.

Option 3: No expansion of deposit insurance, strict liquidity regulation. A simple limit case—though not the global regulatory optimum—is a strict LCR policy that requires that the bank back all its uninsured wholesale deposits with T-bills, so that $B^{Strict} = D_W$ and therefore $L^{Strict} = D_R + E < L^{**} < L^{FB}$. Now there is no moral hazard from expanding deposit insurance, and there are no fire-sale costs (i.e., $h[0] = 0$). The only cost is that with less lending and more bills as assets, the bank forgoes more loans at cost $\pi(L^{FB} - L^{Strict})$.

The basic proposition that follows from this is that if this forgone lending cost is smaller than both the moral hazard cost X and the social costs of the unregulated outcome, then a policy of no deposit insurance for wholesale deposits and a strict T-bill-backed LCR is preferred relative to either the unregulated market outcome or an expansion in deposit insurance. Arguably, our empirical evidence suggests that the costs of forgone lending may be relatively small for most larger banks, specifically that $\pi(L^{FB} - L^{Strict})$ is small.

Of course, the optimally calibrated LCR, which involves lending of $L^{**} > L^{Strict}$ and holding a liquidity buffer of $B^{**} < D_W = B^{Strict}$ T-bills, is always superior to both the unregulated market outcome and the strict LCR. This optimally calibrated LCR will also be superior to a full expansion of deposit insurance if $\pi(L^{FB} - L^{**}) + (1 + \phi)h[L^{**} - D_R - E] < X$.

More generally, one can imagine using various combinations of: (1) more stringent LCR regulation; (2) heightened equity capital requirements; and (3) a partial expansion of deposit insurance to deal with the heightened financial stability risks posed by runs by uninsured wholesale deposits. Indeed, in a richer model, it would arguably make sense to adjust regulatory policy somewhat along all three dimensions—that is, heightened equity capital requirements and a partial expansion of deposit insurance would complement more stringent LCR regulation. Thus, ignoring the political constraints mentioned above, we could envision pairing a more stringent LCR requirement with a modest increase in risk-based equity capital requirements and a targeted expansion of deposit insurance—for example, raising the insurance limit for business payment accounts, one option recently outlined by the FDIC.

LCR DESIGN CONSIDERATIONS Taken at face value, our simple model suggests that a strict LCR requiring full backing of uninsured deposits with T-bills and other short-term Treasuries is preferable to no LCR at all. This would be a dramatic change in the LCR—tantamount to both increasing the runoff rate for uninsured deposits from the current maximum of 40 percent to 100 percent and disallowing all assets that are currently eligible High Quality Liquid Assets (HQLA) except short-term Treasuries and central bank reserves. But even within the scope of the model itself, the optimal policy is something less strict. Moreover, in its simplicity, the model does not speak to all elements of an appropriate regulatory framework. In this section, we propose some considerations relevant to calibrating the strictness of a modified LCR and to specifying the assets that count as HQLA. We then make some qualifications to the simple liquidity assumptions in the model and discuss the relationship of the LCR to the discount window.

At the outset, we note the importance of applying the full LCR to a broader range of banks. As the events in the spring of 2023 demonstrated, there may be contagion from runs even at a midsize regional bank that can endanger a significant part of the banking system. Thus, we strongly favor requiring full LCR compliance by all banks with more than $100 billion in assets, the current statutory threshold for enhanced prudential regulation by the banking agencies. That said, we believe that there is a strong policy case for further lowering the LCR threshold to $50 billion.

Calibrating a modified LCR. It now seems clear that the current maximum runoff rate of 40 percent for uninsured deposits is woefully inadequate. It also seems quite unlikely that it would be socially optimal to require all uninsured deposits to be 100 percent backed by short-term Treasuries. How should the bank regulatory agencies decide where to set the stringency of the LCR between these two boundaries? Starting from the model's implied 100 percent runoff rate, relaxing the strict LCR regulation envisioned in option 3 above may be warranted because not all uninsured deposits are as highly runnable as is assumed in our simple model (even in light of the experience of March 2023), or the costs of restricting socially valuable lending would exceed the financial stability benefits of fully backing uninsured deposits with T-bills, or some combination of the two.

The first justification for relaxation is not reflected in our simple model, which assumes all uninsured deposits to have identical characteristics.

Were regulators to be convinced that some forms of uninsured deposits—such as those used by businesses to meet payrolls and make routine payments to suppliers—were genuinely less prone to run, then the amounts of uninsured deposits to be backed could be reduced.

As discussed earlier, a second justification for relaxing the strict LCR rule—that is, using the optimally calibrated LCR which involves holding T-bills equal to $B^{**} < B^{Strict} = D_W$—arises due to the marginal social costs $\pi > 0$ of reducing lending below the first-best level of L^{FB}. It is admittedly not clear how to translate this concept from the model into a simple metric that can guide the implementation of regulation. However, one factor that should probably be considered is the elasticity of substitution for the loans in question. For example, if a bank cuts back on making on-balance sheet conforming mortgage loans, the marginal social costs π are unlikely to be very high, as these loans can easily find their way into an MBS pool. By contrast, if the marginal loans are opaque to small businesses, finding an alternative provider of credit may involve more friction and hence greater marginal social cost π.

Finally, a third possible rationale for relaxing the strict LCR rule is the concern that an overly strict LCR could have unintended consequences to the extent that it leads to increased money creation activity in the so-called shadow banking system. Concretely, if a strict LCR makes banks more reluctant to take uninsured deposits, investors seeking safe, short-term alternatives may park their cash in money market funds. Flush with cash from savers and facing a shortage of short-term Treasuries (more of which would be owned by LCR-constrained banks), money market funds might conceivably increase their lending against long-term Treasuries and MBS on a short-term collateralized basis through the repo market. The expanded supply of repo financing might in turn raise the incentive of hedge funds and other levered nonbank institutions to finance their long-term securities by borrowing short term.[22]

While all three of these concerns are legitimate, they essentially suggest that a more stringent LCR must be appropriately calibrated to maximize the net benefits, not that the policy direction itself is ill-advised.

Two other points relevant to calibration are worth noting. First, even as one assesses reasons for relaxing the strict LCR rule implied in option 3,

22. This concern may be somewhat mitigated if, as in our example, the increased repo financing is done only against government-backed collateral such as Treasuries and agency MBS. In this case, the potential damage associated with disorderly fire-sale liquidations would seemingly be relatively modest.

there may be other considerations favoring a relatively more stringent requirement. For example, the more severe the fire-sale externalities ϕ, the more stringent should be the LCR. In other words, an increase in ϕ pushes the T-bill holdings B^{**} in the optimally calibrated LCR up toward $B^{Strict} = D_W$. Second, as regulators balance the considerations identified here, they might formulate a more nuanced rule—for example, one alternative would be to progressively increase the assumed outflow rate on a bank's uninsured deposits as its uninsured deposits rise as a share of its total deposits.

Eligible HQLA. A second important consideration in designing a revised LCR is the definition of HQLA—both the specification of assets that qualify and any limitations or conditions in counting them against runnable liabilities. The same calibrated run rate for uninsured deposits will have quite different impacts upon banks depending on the range of assets that qualify as HQLA. Thus, another way to effectively relax the strict LCR contemplated in option 3 is by allowing the bank to meet some or all of its requirement to back uninsured deposits with all assets that qualify as HQLA under the current LCR, rather than just short-term Treasuries. The most important consideration here is whether there should be any change in the eligibility of long-term securities such as ten-year Treasuries and agency-backed MBS. As we have seen, within their holdings of liquid assets, banks have a very strong preference for longer-duration securities.[23]

At present, longer-duration Treasuries count as unlimited HQLA, based on current market value, while agency-backed MBS may count for up to 40 percent of total HQLA, with a 15 percent haircut off current market value. However, from a social perspective, longer-duration securities are an inefficient way to back uninsured deposits. This is because longer-duration securities, even if they remain completely liquid, may have a lower market value in the bad state. Thus, a bank would have to hold $1/F_S$ units of long-term securities, rather than just one unit of T-bills, to prevent the same

23. Under the current LCR, HQLA are divided into level 1, level 2A, and level 2B assets. Level 1 assets consist of all US Treasuries, reserves, other liquid obligations fully backed by the US government, and liquid obligations of very low-risk foreign sovereigns and international institutions. Level 2A assets consist of agency-backed MBS, other agency-backed debt, and liquid obligations of low-risk foreign sovereigns. Level 2B assets consist of investment-grade nonfinancial corporate bonds, investment grade municipal bonds, and large-cap US public equities. Irrespective of their maturities, level 1 assets are subject to a 0 percent haircut, while level 2A and 2B assets are subject to haircuts of 15 percent and 50 percent, respectively. Furthermore, level 2B assets cannot account for more than 15 percent of a bank's total HQLA; and the sum of level 2A and 2B assets cannot account for more than 40 percent of HQLA.

amount of socially inefficient fire-selling of loans. This in turn would crowd out more valuable lending ex ante, with no social benefit, since society is not obviously better served by having banks hold long-term securities as opposed to T-bills, even if bankers privately prefer the former.[24] Again, an important point here is that, from a social perspective, the intermediation of long-term securities can be more safely done in the bond fund sector, where investors knowingly assume the interest rate risk themselves, than with runnable uninsured bank deposits.

On the other hand, it is unrealistic to think that all banks could back their currently high levels of uninsured deposits with short-term Treasuries and reserves alone. To put this issue in perspective, there is currently about $8.3 trillion of outstanding Treasury debt that matures within the next twelve months (this includes $5.7 trillion of T-bills and $2.6 trillion of short-term notes and bonds), along with about $3.5 trillion of reserves (a figure that is diminishing by about $80 billion a month as the Federal Reserve continues its program of quantitative tightening). There are about $8 trillion of uninsured deposits.[25] So an average assumed runoff rate of 75 percent for uninsured deposits would require using more than half of all reserves and outstanding short-term Treasuries as backing, while an assumed runoff rate of 100 percent would consume about two-thirds of those two asset classes. Thus, as a practical matter, there is reason to allow longer-duration securities that carry essentially no credit risk.

To be clear, this simple calculation ignores equilibrium effects. In particular, imposing a more stringent form of the LCR on uninsured deposits will reduce the quantity of uninsured deposits in the system, which we

24. One reason bankers might have a private preference for long-term securities is that they have a term premium, which generates higher reported income (Hanson and Stein 2015). To the extent that such a term premium is just compensation for risk, long-term securities are not a socially higher-NPV investment than short-term bills, but they may be attractive to managers whose incentives are to maximize reported earnings. Similarly, bankers might have private preference for MBS over like-duration Treasuries because MBS yields contain an extra option premium component that compensates holders for the fact that they are short a call option on interest rates.

25. US Department of Treasury, "Most Recent Quarterly Refunding Documents," quarterly release data, https://home.treasury.gov/policy-issues/financing-the-government/quarterly-refunding/most-recent-quarterly-refunding-documents; Federal Reserve Board, "Liabilities and Capital: Other Factors Draining Reserve Balances: Reserve Balances with Federal Reserve Banks: Week Average," series WRESBAL, retrieved from FRED, https://fred.stlouisfed.org/series/WRESBAL; FDIC, "FDIC Quarterly Banking Profile," balance sheet, https://www.fdic.gov/quarterly-banking-profile.

view as an entirely desirable outcome, especially to the extent that these deposits are funding long-term securities holdings. Moreover, as noted above, even holding fixed the quantity of uninsured deposits, there is a policy case for offsetting to some degree banks' incentive to back them with longer-duration securities. Thus, the banking agencies might want to consider tightening the current LCR limit of 40 percent that applies to agency MBS and imposing some form of limit on the portion of longer-term Treasury securities that can count as HQLA. Alternatively, a similar outcome might be achieved by subjecting eligible longer-term securities to a haircut that steeply increases with the duration of these securities.

Relationship of the modified LCR to the discount window. Our model assumes perfect liquidity for both T-bills and longer-duration securities. However, as observed during both 2008 and 2020, the immediate liquidity of even the safest assets can have limits during periods of serious financial dislocation. Moreover, as was evidenced during the bank panics in the spring of 2023, practical impediments such as the need to move collateral may stymie banks' attempts to access the discount window quickly when other avenues of funding have been closed off. For both these reasons, we believe that any required backing of uninsured deposits under a modified LCR, including T-bills, should be pre-positioned at the discount window.

With or without a requirement for pre-positioning, the question arises whether loans pre-positioned at the discount window should be credited for purposes of satisfying the LCR—both generally and for backing uninsured deposits in the kind of regime we propose. Of course, loans on the books of banks do not qualify as HQLA under the current LCR. But another way for a bank to generate liquidity at time 1—and hence to avoid fire sales of its loans—is to borrow from the discount window using these loans as collateral. In fact, as part of their liquidity management strategies, some banks already pre-position significant portions of their loan portfolios at the discount window. Thus, one might argue that the LCR should give banks credit for this lender of last resort (LOLR) access if they are willing to pre-position the loan collateral at time 0 and allow it to serve, in addition to T-bill holdings, as backing for uninsured deposits. Indeed, a recent report by the Group of Thirty (2024) makes just that recommendation.

To consider this possibility, we assume that the Federal Reserve, as LOLR, is restricted to making loans at time 1 that are fully collateralized, that is, loans that are virtually certain to be fully repaid at time 2. If not, it would be taking nontrivial credit risk, something that it is not legally authorized to do through its discount window lending under section 10B of the Federal Reserve Act.[26] Accordingly, if a bank pre-positions loans in amount L at the discount window at time zero, it can count on being able to borrow only $z_L L$ at time 1, where again, z_L is the worst-case value of the loans at time 2. Thus far we have assumed that $z_L = 0$, implying that banks cannot use loans to collateralize any discount window borrowing at time 1.

However, the model can be easily extended to cover the case where $z_L > 0$ so that loans can be used to collateralize borrowing from the Fed at time 1. The analysis of option 1 (above) is identical to the case where $z_L = 0$. There is no need for LOLR borrowing at time 1, because all deposits are insured, and hence there are no runs. In option 2, as long as $D_W > z_L L^* + F_S S^* + B^*$, the bank will be unable to fully pay off departing uninsured depositors just by selling its liquid assets and borrowing against its loans at the discount window. Rather, it will now have to liquidate a fraction Δ_L of its loans so that $\Delta_L k_L F_L L^* + (1 - \Delta_L) z_L L^* + F_S S^* + B^* = D_W$. In other words, the LOLR policy reduces the amount of fire-selling (i.e., Δ_L is now smaller all else being equal), because some liquidity is obtained from the LOLR at time 1.

Similarly, in the strict LCR of option 3, the bank does not need to hold as many bills as before in order to completely avoid fire sales. Now we only require that $z_L L + F_S S + B = D_W$. This allows for more lending ex ante and yet still satisfies the requirement that the combination of liquid assets and discount window access be enough to pay off all uninsured depositors in the event of a run at time 1, without having to inefficiently liquidate any loans at this date.

Thus, the model suggests that, subject to appropriate collateral haircuts, it may be sensible to allow loans that are pre-positioned at the discount window to count toward satisfying an LCR for uninsured deposits. Doing so would accord with the aim of ensuring that the LCR does not overly constrain banks' ability to use uninsured deposits to finance positive-NPV loans. Still, it is important to recognize that the issues associated with setting an appropriate haircut on pre-positioned loan collateral in the context of a regulatory requirement would be very different, and considerably thornier, than those that arise in traditional discount window operations.

First is the question of the time horizon. If a bank approaches the discount window ex post, at the moment it needs to borrow, the haircut on

26. Federal Reserve Board, "Federal Reserve Act: Section 10B. Advances to Individual Member Banks," https://www.federalreserve.gov/aboutthefed/section10b.htm. Section 10B requires that any advances to member banks be "secured to the satisfaction" of the Reserve bank making the advance. As reflected in the Federal Reserve's policies on discount window lending, this provision is understood to require sufficient collateralization to virtually guarantee that the Reserve bank will be repaid in full.

collateral is set at the time the loan is extended. If, as is presently the case, a bank chooses to pre-position loans as a precautionary measure, it is doing so as part of its own business strategy. Here, by contrast, we are contemplating a situation where a bank is given ex ante regulatory credit for discount window borrowing that it might undertake at some later date, months or even years into the future. At this longer horizon, there is obviously a greater risk that the collateral will decline in value. In the language of the model, this is tantamount to saying that z_L is likely to be far below 1.

Indeed, it is possible that the prospect of a run is either prompted by concerns about the quality of a bank's loans or, even if a run is set off by other reasons, reveals that its loan book has been opaquely declining in value. In these circumstances, the ordinary response of requiring more collateral to compensate for the decline in value of existing collateral could exacerbate the already deteriorating liquidity situation of the bank. Alternatively, were the Federal Reserve to continue to promise availability at the original value of the loan collateral, it would effectively be taking on credit risk. Thus, haircuts for loans would have to be set more conservatively for LCR purposes.[27]

Second, the logistics of a regime in which allowing pre-positioned loans to meet LCR requirements could be daunting. Precisely because there are no readily identifiable market values for loans, as there are for traded securities, the Federal Reserve's schedule of collateral haircuts has very wide ranges for each category of loan.[28] The actual haircut imposed for any individual loan is determined by a model maintained by the Federal Reserve Bank of New York. Were loans pre-positioned at the discount window to be treated as HQLA, the complexity of this process might have to increase dramatically, with consequent increased risks of mistakes. Regular revaluation of all pre-positioned loan collateral by banks taking advantage of this new form of HQLA would, if taken seriously, be potentially much more burdensome—and imprecise—than repricing securities with observable market values. In this sense, the qualitative argument in favor of a largely T-bill-backed LCR remains similar to that above.

Finally, the inherent imprecision of setting haircuts at such longer horizons, combined with the heightened regulatory stakes at play, suggests that

27. It is important to note that this problem is not fully addressed by the Federal Reserve's current practice of repricing loans pre-positioned at the discount window on a monthly basis, presumably in calm circumstances when a bank could add more collateral. The problem of unknown or hidden losses would remain.

28. For example, the haircut for a commercial real estate loan ranges from 44 percent to 95 percent of its estimated market value.

such a process may give rise to a great deal of lobbying and political pressure around what the appropriate value of the haircuts should be for various types of loans. In short, were the regulatory agencies to go down the road of counting pre-positioned loans as HQLA, we would urge them to proceed cautiously. They might, for example, begin on a relatively small scale—say by creating a new category 2C form of HQLA that would be limited to a small percentage of total HQLA requirements. Over time, if experience with the valuation process gave confidence that a higher limit was prudent, an adjustment could be made.

V.B. Interest Rate Risk and Capital Regulation

In the above discussion, we have taken bank equity capital as exogenously fixed and focused exclusively on liquidity regulation. One conclusion has been that a well-designed LCR should lean against the use of long-duration securities as backing for uninsured deposits. Of course, interest rate risk can also be addressed with capital requirements. The current risk-based capital regime does not do this for the banking book.[29] In fact, the US banking agencies have only partially implemented the framework for supervisory oversight of bank management of interest rate risk originally developed by the Basel Committee on Banking Supervision in 2004 and updated in 2016. Remarkably, the Federal Reserve's stress test scenarios in 2021 and 2022 did not include interest rate increases—something most observers would have identified as an obvious risk to the industry at that time. Even without the broad evolutionary changes to the banking industry that we have highlighted, a more rigorous and complete coverage of interest rate risk in capital requirements would seem warranted. Those changes, though, considerably strengthen the case. As was painfully apparent in March 2023, large portfolios of longer-duration debt securities can meaningfully increase banks' vulnerability to significant changes in market interest rates.

Moreover, interest rate risk on the asset side interacts in an important way with factors that make deposits more likely to either reprice or run. Conventional wisdom has held that interest rate risk in the banking book was to some extent hedged by the stickiness of deposits. That is, although interest rate hikes reduced the present value of a bank's assets, this decline in asset value was offset by an increase in the value of the deposit franchise to the extent that the bank could retain most of its deposits, even if it increased the interest rate it paid on these deposits by only a fraction of the central bank's

29. Interest rate risk is considered in calculating risk-weighted requirements for the trading books of large banks.

target rate increase. But if the deposit beta has increased, pressure on bank earnings and, eventually, capital may build more quickly.[30] See Drechsler and others (2023) for a recent analysis along these lines.

Going further, banks other than the very largest have not even been required to recognize unrealized changes in market value of their securities holdings—for example, due to a rise in interest rates—in their regulatory capital metrics. This is due to hold-to-maturity accounting and the accumulated other comprehensive income (AOCI) opt-out election for securities that are accounted for on an available-for-sale (AFS) basis. The banking agencies have now proposed to eliminate this AOCI opt-out for banks with assets between $100 billion and $700 billion (OCC, Federal Reserve Board, and FDIC 2023b). If this regulatory change is adopted, mark-to-market gains and losses on AFS securities will begin to have an impact on the reported regulatory capital of midsize regional banks. We view this as a useful step in addressing interest rate risk, though it would probably be preferrable to have an explicit capital requirement for duration risk in banking book securities portfolios. Additionally, the regulatory agencies must decide how to treat securities designated as hold-to-maturity. It is unclear to what extent a change in rules applied only to the AFS book might be gamed by banks reclassifying AFS securities as hold-to-maturity.[31]

30. The deposit beta is a measure of the sensitivity of the interest expense on a bank's deposits to changes in short-term money market rates (e.g., the federal funds rate).

31. Banks account for their securities in three different ways under US generally accepted accounting principles (GAAP). Trading account securities are carried on the balance sheet at their current market value, so any mark-to-market gains and losses have an impact on book equity and flow through net income. Securities a bank intends to hold until maturity are recorded in the hold-to-maturity account and are carried at their historical amortized cost. Fluctuations in the market value of hold-to-maturity securities due to changes in level of interest rates do not have an impact on the bank's book equity or its net income. Securities a bank might sell prior to maturity are recorded in the AFS account. AFS securities are carried at their market value and fluctuations in mark-to-market value of AFS securities have an impact on book equity. However, unrealized mark-to-market gains and losses on AFS securities do not affect net income and the retained earnings equity account. Instead, these mark-to-market changes are recorded in a different equity account—the AOCI—and are only recognized in net income if the bank sells the security. While unrealized fluctuations in the mark-to-market value of AFS securities have an impact on accounting book equity, the AOCI opt-out refers to the fact that, since 2013, US bank regulators have allowed banks other than the very largest to ignore mark-to-market changes in the value of AFS when computing their regulatory equity capital. This means that, while they differ for GAAP purposes, there is almost no difference between AFS and hold-to-maturity securities from the standpoint of regulatory capital. The very largest G-SIB banks already must pass through to capital any changes in the market value of their AFS securities. Until the Federal Reserve's 2019 tailoring regulation, *all* banks with over $250 billion were also required to do so.

Finally, unless the Federal Reserve's annual supervisory stress test is again applied to all banks over $50 billion, as was the case before a legislative change in 2018, even a regular stress test scenario focused on interest rate risk would miss many vulnerable banks.[32] Thus, while we are aware of the prevailing view of regulators that a more generally applicable interest rate risk rule is infeasible, we believe that the regulators should try again. If the effort proves unsuccessful, a second-best approach would be a structured supervisory program that regularly assessed the interest rate risk of all banks above a certain size threshold.

V.C. Merger and Competition Policy

Our analysis supports the view that changes in the industry have threatened the business model of many midsize regional banks. As such, our analysis has implications for bank merger policy, as well as prudential regulation.

Midsize banks risk being caught between the scale economies of the largest banks and the relationship-lending capabilities of community banks. Increasing returns to scale have already been achieved in most forms of consumer lending through the standardization of credit analysis. In recent years, scale has also allowed the largest banks to invest substantial amounts in information technologies. As algorithms become more sophisticated and artificial intelligence enters credit decision making, size will likely be further rewarded since there are significant economies of scale in these sorts of IT investments. At the same time, community banks are likely better positioned to take advantage of the remaining opportunities for relationship lending and payoffs to localized knowledge, notably in lending to smaller businesses.

This characterization of the industry is reinforced by the trends we have identified—notably the changes in the portfolios of midsize regional banks, with the decline in C&I lending and the increase in securities holdings. These changes may leave these midsize banks in the uncomfortable position of having to rely very heavily on their deposit franchise—that is, the ability to pay submarket rates to their depositors—for a disproportionate share of their value creation. As was demonstrated in March 2023, the franchise value of this group of banks has likely been further eroded by the increasing ease and speed with which deposits can be moved across banks.

32. Economic Growth, Regulatory Relief, and Consumer Protection Act, Pub. L. 115–174, 132 Stat. 1296 (May 24, 2018), https://www.congress.gov/bill/115th-congress/senate-bill/2155.

If our assessment is on target, the economics of the industry may lead to a significant deterioration in the competitive position of midsize regional banks in the coming years. How this plays out will depend in significant part on the regulatory response. Will the banking agencies allow the capital, liquidity, and earnings positions of a set of increasingly uncompetitive banks to deteriorate? As the savings and loan crisis of the 1980s and early 1990s showed, such forbearance can end up being very costly for the economy and taxpayers. If, instead, the agencies maintain or increase regulatory rigor to prevent these banks from taking excessive risks in a desperate search for profits, then they may just stagnate. In that scenario, the business they lose will probably be captured by larger banks. The result would be a further increase in concentration of the banking industry.

In the face of these possibilities, it may be wise for bank merger policy to acknowledge these competitive dynamics and to look more positively on mergers of midsize regional banks and on acquisitions of smaller banks by regional ones. It is hard to say whether these combinations will be able to achieve the scale economies needed for these banks to thrive over the long run. But at least they would create institutions that are better able to compete with the largest banks. While a strict antitrust policy for the mega-banks is entirely reasonable, a similarly strict policy for the midsize regional banks might—ironically—redound to the benefit of those same mega-banks.

VI. Conclusions

Our review of bank balance sheets over the last quarter century shows that, while uninsured deposits have become a greater share of liabilities, the information-intensive lending that dominates traditional views of banking has declined as a share of assets. While these trends on the deposit side might stall or reverse, the fact that they predated the Federal Reserve's responses to the global financial crisis suggests that the rapid growth in deposits—and the rising share of those deposits that are uninsured—are developments warranting attention from regulators. Similarly, there are good reasons to believe that migration of business lending to nonbank institutions—especially lending to large and medium-size businesses—is likely to continue unabated.

One insight that emerges from the confluence of these two trends is that regulators may be more comfortable tightening liquidity requirements on uninsured deposits, given that the substantial increase in those deposits in recent decades has not been correlated with an increase in information-intensive lending. On the contrary, the two appear to be negatively correlated. A second conclusion is that the regulation of midsize regional banks may be especially in need of attention. As noted, the business model of these banks looks increasingly vulnerable. At the same time, unlike the G-SIBs and the very largest regional banks, these banks are not currently subject to regulation and supervisory programs that account for the increased runnability of deposits.

Our effort here has been to provide some foundation for fashioning appropriate regulatory responses and some considerations to bear in mind in doing so. More work will obviously need to be done by researchers and regulators to calibrate and build out specific proposals.

ACKNOWLEDGMENTS We thank the editors Jan Eberly and Jón Steinsson, our discussants Arvind Krishnamurthy and Raghuram Rajan, as well as Bill English for extremely helpful comments.

References

Acharya, Viral V., Rahul S. Chauhan, Raghuram G. Rajan, and Sascha Steffen. 2024. "Liquidity Dependence and the Waxing and Waning of Central Bank Balance Sheets." Working Paper 31050. Cambridge, Mass.: National Bureau of Economic Research.

Acharya, Viral V., and Raghuram G. Rajan. 2023. "Liquidity, Liquidity Everywhere, Not a Drop to Use: Why Flooding Banks with Central Bank Reserves May Not Expand Liquidity." Working Paper 29680. Cambridge, Mass.: National Bureau of Economic Research. (Forthcoming in *Journal of Finance.*)

Admati, Anat R., Peter M. DeMarzo, Martin F. Hellwig, and Paul Pfleiderer. 2013. "Fallacies, Irrelevant Facts, and Myths in the Discussion of Capital Regulation: Why Bank Equity Is Not Expensive." Working Paper 161. Stanford, Calif.: Rock Center for Corporate Governance at Stanford University.

Bates, Thomas W., Kathleen M. Kahle, and Rene M. Stulz. 2009. "Why Do U.S. Firms Hold So Much More Cash Than They Used To?" *Journal of Finance* 64, no. 5: 1985–2021.

Benmelech, Efraim, Jun Yang, and Michal Zator. 2023. "Bank Branch Density and Bank Runs." Working Paper 31462. Cambridge, Mass.: National Bureau of Economic Research.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2024. "The Secular Decline of Bank Balance Sheet Lending." Working Paper 32176. Cambridge, Mass.: National Bureau of Economic Research.

Cookson, J. Anthony, Corbin Fox, Javier Gil-Bazo, Juan Felipe Imbet, and Christoph Schiller. 2023. "Social Media as a Bank Run Catalyst." Working Paper 4422754. Paris: Université Paris-Dauphine.

Diamond, Douglas W. 1984. "Financial Intermediation and Delegated Monitoring." *Review of Economic Studies* 51, no. 3: 393–414.

Diamond, Douglas W., and Philip H. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91, no. 3: 401–19.

Diamond, Douglas W., and Raghuram G. Rajan. 2001. "Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking." *Journal of Political Economy* 109, no. 2: 287–327.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl. 2017. "The Deposits Channel of Monetary Policy." *Quarterly Journal of Economics* 132, no. 4: 1819–76.

Drechsler, Itamar, Alexi Savov, Philipp Schnabl, and Olivier Wang. 2023. "Banking on Uninsured Deposits." Working Paper 31138. Cambridge, Mass.: National Bureau of Economic Research.

Egan, Mark, Stefan Lewellen, and Adi Sunderam. 2022. "The Cross-Section of Bank Value." *Review of Financial Studies* 35, no. 5: 2101–43.

Erel, Isil, and Eduard Inozemtsev. 2024. "Evolution of Debt Financing toward Less-Regulated Financial Intermediaries in the United States." Working Paper 32114. Cambridge, Mass.: National Bureau of Economic Research.

Federal Deposit Insurance Corporation (FDIC). 2023. *Options for Deposit Insurance Reform*. Arlington, Va.: Author. https://www.fdic.gov/analysis/options-deposit-insurance-reforms/index.html.

Gatev, Evan, and Philip E. Strahan. 2006. "Banks' Advantage in Hedging Liquidity Risk: Theory and Evidence from the Commercial Paper Market." *Journal of Finance* 61, no. 2: 867–92.

Gorton, Gary, and George Pennacchi. 1990. "Financial Intermediaries and Liquidity Creation." *Journal of Finance* 45, no. 1: 49–71.

Graham, John R., and Mark T. Leary. 2018. "The Evolution of Corporate Cash." *Review of Financial Studies* 31, no. 11: 4288–344.

Group of Thirty. 2024. *Bank Failures and Contagion: Lender of Last Resort, Liquidity, and Risk Management*. Washington: Author. https://group30.org/publications/detail/5264.

Hanson, Samuel G., Anil K. Kashyap, and Jeremy C. Stein. 2011. "A Macroprudential Approach to Financial Regulation." *Journal of Economic Perspectives* 25, no. 1: 3–28.

Hanson, Samuel G., Andrei Shleifer, Jeremy C. Stein, and Robert W. Vishny. 2015. "Banks as Patient Fixed-Income Investors." *Journal of Financial Economics* 117, no. 3: 449–69.

Hanson, Samuel G., and Jeremy C. Stein. 2015. "Monetary Policy and Long-Term Real Rates." *Journal of Financial Economics* 115, no. 3: 429–48.

Kashyap, Anil K., Raghuram G. Rajan, and Jeremy C. Stein. 2002. "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking." *Journal of Finance* 57, no. 1: 33–73.

Koont, Naz, Tano Santos, and Luigi Zingales. 2023. "Destabilizing Digital 'Bank Walks.'" Working Paper 328. Chicago: George J. Stigler Center for the Study of the Economy and the State.

Lerner, Josh, Amit Seru, Nicholas Short, and Yuan Sun. 2024. "Financial Innovation in the Twenty-First Century: Evidence from US Patents." *Journal of Political Economy* 132, no. 5: 1391–449.

Li, Lei, Philip E. Strahan, and Song Zhang. 2020. "Banks as Lenders of First Resort: Evidence from the COVID-19 Crisis." *Review of Corporate Finance Studies* 9, no. 3: 472–500.

Lopez-Salido, David, and Annette Vissing-Jorgensen. 2023. "Reserve Demand, Interest Rate Control, and Quantitative Tightening." Working Paper. https://www.ecb.europa.eu/press/conferences/shared/pdf/20231004_mon_pol_conference/Lopez_Salido_paper.pdf.

Office of Inspector General (OIG). 2023. *Material Loss Review of Silicon Valley Bank.* Evaluation Report 2023-SR-B-013. Washington: Board of Governors of the Federal Reserve System and Consumer Financial Protection Bureau (CFPB). https://oig.federalreserve.gov/reports/board-material-loss-review-silicon-valley-bank-sep2023.pdf.

Office of the Comptroller of the Currency (OCC), Board of Governors of the Federal Reserve System (Federal Reserve Board), and Federal Deposit Insurance

Corporation (FDIC). 2023a. *Shared National Credit Program: 1st and 3rd Quarter 2022 Reviews.* Washington: Authors. https://www.fdic.gov/sites/default/files/2024-03/pr23012a.pdf.

Office of the Comptroller of the Currency (OCC), Board of Governors of the Federal Reserve System (Federal Reserve Board), and Federal Deposit Insurance Corporation (FDIC). 2023b. *Interagency Overview of the Notice of Proposed Rulemaking for Amendments to the Regulatory Capital Rule: Applicable to Large Banking Organizations and to Banking Organizations with Significant Trading Activity.* Washington: Authors. https://www.fdic.gov/sites/default/files/2024-03/basel-3-reforms-npr-reference-guide-07-27-23_0.pdf.

Schneider, Thomas, Philip E. Strahan, and Jun Yang. 2023. "Banking in the United States." Manuscript. https://drive.google.com/file/d/18dg1hRG6NL6bUaCy45aI9IXRaxen2p9f/view?pli=1. (Forthcoming in *Oxford Handbook of Banking*, 4th edition, edited by Allen N. Berger, Philip Molyneux, and John O. S. Wilson. Oxford: Oxford University Press.)

Stulz, René M., Alvaro G. Taboada, and Mathijs A. van Dijk. 2024. "Why Are Banks' Holdings of Liquid Assets So High?" Working Paper 30340. Cambridge, Mass.: National Bureau of Economic Research.

Wheat, Chris, and Erica Deadman. 2023. "Household Pulse: Balances through March 2023." JPMorgan Chase Institute. https://www.jpmorganchase.com/institute/all-topics/financial-health-wealth-creation/household-pulse-cash-balances-through-March-2023.

Comments and Discussion

COMMENT BY

ARVIND KRISHNAMURTHY US regional banks experienced financial stress in the spring of 2023. In the case of Silicon Valley Bank (SVB), this stress led to a bank run and ultimate failure. The mix of ingredients that drove stress in SVB are by now well documented: a large fraction of uninsured deposits, investments in long-duration securities that suffered losses as interest rates rose, and poor risk management (Barr 2023; Jiang and others 2023).

As we pass the one-year anniversary of the SVB failure, should we be sanguine about the state of the banking system? The authors make a persuasive case that we should not. The mix of ingredients that drove the banking stresses has been years in the making. The authors take a longer view of the evolution in banking and highlight three key trends.

First, the quantity of deposits relative to GDP has risen across the banking system, and the share of uninsured to total deposits has also risen.

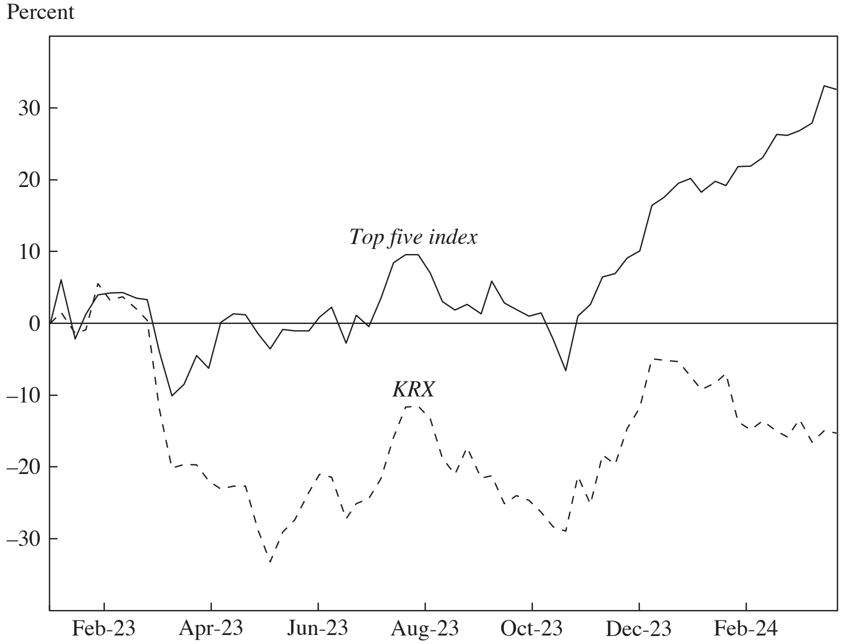
Second, many of the loan-making activities of banks have migrated to nonbanks. These loan-making activities, which require information such as screening or monitoring, are now being performed as effectively, if not more so, by nonbanks. Moreover, this trend has accelerated in the last decade.[1]

Third, in response, the largest banks have shifted toward a liquidity provision model. They offer deposits to customers, holding these deposits in securities such as Treasuries or mortgage-backed securities. They also offer credit lines, a form of contingent liquidity provision, to their corporate customers. These activities are currently not being performed by nonbanks.

1. Buchak and others (2024) document the diminishing role of banks in lending and show that it can have important implications for the monetary transmission mechanism.

Brookings Papers on Economic Activity, Spring 2024: 390–411

Figure 1. Top Five Bank Index versus KBW Regional Bank Index (KRX)



Source: LSEG Data & Analytics.

The regional banks, which previously were active in information-sensitive lending, have had their business eroded. Some, such as SVB, have shifted to providing uninsured deposits to customers and holding securities to back these deposits. But the entry of nonbanks creates risk for their business model.

Figure 1 compares the cumulative stock return on a value-weighted index of the five largest US banks (JPMorgan Chase, Citigroup, Bank of America, Wells Fargo, Goldman Sachs) to that of the KBW Nasdaq Regional Banking Index (KRX). Prior to March 2023, these indexes tracked each other. The banking stresses last spring led to a fall in regional bank stock prices and a rise in the large bank stock prices. By the end of the sample, this divergence has accumulated to a 47.9 percent relative return between the large and midsize banks. The assessment of the authors—that the regional bank model is under stress—is evident in the figure.

The authors then offer three policy proposals to guard against the stresses in the banking system. First, they propose a tightening of the liquidity coverage ratio. Second, they propose that interest rate risk from long-duration securities be subject to risk-based capital requirements and that

Figure 2. Balance Sheet Model of Bank

Assets	Liabilities
Loans (L)	Deposits (D)
Tradable securities (S)	*(Book) Equity*
Tangible assets (A)	Liabilities and equity

Source: Author's illustration.

gains and losses from such security holdings pass through to regulatory measures of bank capital. Finally, they recommend that the government adopt a receptive stance toward bank mergers in the midsize regional banking sector.

In my comments, I will focus on these policy proposals, evaluating them and offering suggestions to strengthen each one. Overall, I strongly endorse the proposals put forward in the paper. I share the authors' view that regional banks are in a financially precarious position that is masked because investors currently assess that their uninsured deposits are effectively backstopped by the government. I also think that action is urgently needed, and that not doing so risks kicking the can down the road.

BALANCE SHEET MODEL OF A BANK Consider the following model of a bank (figure 2). The bank raises funds through deposits and issuing equity. These funds are used to make loans or hold tradable securities. The bank's security purchases, along with its equity issuance, are all market-based transactions with zero net present value at the time of trade.

The bank can create value via its deposit-taking and lending activities. This value is reflected in the interest rate spread the bank offers on deposits and charges on loans, relative to the equivalent market rate. These spreads can exist because of the bank's market power, informational advantages, provision of transaction services, and so on. Define the interest rate spread on deposits and loans, relative to the short-term funding rate, as follows:

$$\text{Deposit rate spread} \equiv r^* - r^D, \text{ Loan rate spread} \equiv r^L - r^*,$$

where r^D is the average rate paid on deposits, r^L is the average rate earned on loans, and r^* is the short-term market interest rate (e.g., the federal funds rate).

Given D as the total amount of deposits, and L as the total amount of loans, the total cash flow generated by these rate spreads is given by:

$$R = D\left(r^* - r^D\right) + L\left(r^L - r^*\right).$$

Suppose that the bank incurs a per period cost of C to operate and earn these spreads. Then the value of the bank franchise is the present value of the net cash flow, $PV(R - C)$.[2]

BANK THRESHOLDS A bank has two important financial thresholds, one governing solvency and the other liquidity. We can compute,

$$Market\ Equity = \left(L + S - D\right) + MTM_{L,S} + PV\left(R - C\right).$$

The market value of equity is the sum of the assets minus deposits, with an adjustment for any mark-to-market gains or losses on the loans and securities ($MTM_{L,S}$), and the present value of the bank franchise. A bank is *solvent* if this is positive. Capital requirements key off solvency.

A bank is *liquid* if the cash that can be raised from loans and securities covers all of its deposits:

$$\left[L - hL\right] + S - D + MTM_{L,S} > 0.$$

Here h is the haircut on loans, and assume there is a zero haircut on securities. The haircut reflects the fire-sale loss that comes from selling assets. Relative to the solvency threshold, liquidity does not include franchise value and includes a haircut on loans.

The liquidity coverage ratio (LCR) is a requirement on the liquidity threshold. Typically, only securities are considered as available liquidity for regulatory purposes and the LCR requirement is:

$$S + MTM_{L,S} - \lambda D > 0,$$

where λ is an assumed runoff rate on deposits. In current bank regulation, this runoff rate is 40 percent on uninsured deposits.

The authors propose three changes to the LCR. First, they propose to increase the runoff rate above 40 percent. Given the speed of the bank run at SVB, a move to increase the runoff rate is warranted. This increase is further justified when considering the broader fact that nonbanks are a substitute for banks in loan making. That is, the social cost of tightening the LCR is that it crowds out lending by banks, but this is less socially costly to the extent that there are good substitutes for bank credit.

2. See DeMarzo, Krishnamurthy, and Nagel (2024) and Drechsler and others (2023) for an analysis of the bank's franchise value.

Third, the authors propose that banks pre-position the securities used to satisfy the LCR at the discount window. I also endorse this proposal, which has been made by others (Duffie 2024; Group of Thirty 2024; Hsu 2024). I see the rationale as primarily operational. In practice, banks turn to the Federal Home Loan Banks (FHLBs) for liquidity during a crisis, rather than the discount window. This appears to happen because FHLBs offer liquidity cheaper than the discount window and because the discount window may create some stigma. In other banking systems, the discount window is the key source of liquidity in a crisis. I see it as low-hanging fruit—no cost and only benefit—if their proposal catalyzes the discount window to operate as intended in a crisis.

DURATION RISK As noted above, the authors propose that the interest rate risk on securities be recognized both in computing capital requirements and for bank accounting. Given the failure of interest rate risk management in the banking crisis, this too is warranted.

Duration considerations also enter in the choice of what set of securities can be used to meet the LCR. The authors propose that reserves and Treasury bills (T-bills) be used, concerned that long-term Treasuries may have low value in a crisis event. But, as the authors note, there are equilibrium issues that arise in this case: "assumed runoff rate of 75 percent for uninsured deposits would require using more than half of all reserves and outstanding short-term Treasuries as backing, while an assumed runoff rate of 100 percent would consume around two-thirds of those two asset classes." In equilibrium, such a proposal would depress T-bill yields and may distort issuance decisions. For example, it would incentivize the US Treasury to shorten issuance maturity, which may increase fiscal risk.

I propose investigating another option, but one which would require the discount window to catch up to modern securities markets. Much interest rate risk is managed using interest rate swaps. These swaps are now plain vanilla, with standardized collateral arrangements. A long-duration Treasury bond plus an interest rate swap to hedge the duration risk is equivalent in risk terms to a short-duration Treasury. Sophisticated banks trade in both securities and interest rate swaps regularly. Thus, I propose that the LCR be satisfied by the combination of a long-term Treasury and swap, and that this package also be pre-positioned at the discount window as collateral.

LIQUIDITY RISK AND BANK CAPITAL An important observation in the SVB episode is that a liquidity problem, even if it is eased by liquidity from the government, can turn into a solvency problem. Consider the case where some uninsured depositors, say corporate business clients, withdraw their deposits from a bank they are nervous about. The bank then turns to the discount window to source the liquidity to pay these depositors. But in the process, the bank replaces a profitable source of deposits costing r^D with a discount loan at a rate greater than r^*. As a result, the franchise value of the bank $PV(R - C)$ falls. Thus, losing business (the corporate depositor) erodes franchise value and the liquidity problem becomes a solvency problem. Indeed, if the bank is not well capitalized, the financial stress will worsen, and the bank will be forced to close.

Another way of stating my point is to note that the franchise value of a bank, $PV(R - C)$, is a risky bank asset, where the risk arises from liquidity concerns.

An immediate implication is that capital requirements should be *liquidity-based* and not just *risk-based* as in current practice. Thus, I would further propose that capital requirements be strengthened in this manner (DeMarzo, Krishnamurthy, and Nagel 2024; DeMarzo and others 2023).

CONCLUSION The regional bank model is under stress. Uninsured deposits are high in aggregate and in particular pockets. The authors propose a tighter LCR in the face of flighty uninsured deposits, capital charges on interest rate risk, and pre-positioning collateral at the discount window. I strongly endorse these proposals. I would also go further, particularly in terms of tightening capital requirements, linked to liquidity risk.

REFERENCES FOR THE KRISHNAMURTHY COMMENT

Barr, Michael S. 2023. *Review of the Federal Reserve's Supervision and Regulation of Silicon Valley Bank*. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/publications/files/svb-review-20230428.pdf.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2024. "Beyond the Balance Sheet Model of Banking: Implications for Bank Regulation and Monetary Policy." *Journal of Political Economy* 132, no. 2: 616–93.

DeMarzo, Peter, Erica Xuewei Jiang, Arvind Krishnamurthy, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2023. "Resolving the Banking Crisis." *ProMarket*, April 20. https://www.promarket.org/2023/04/20/resolving-the-banking-crisis/.

DeMarzo, Peter, Arvind Krishnamurthy, and Stefan Nagel. 2024. "Interest Rate Risk in Banking." Working Paper 4194. Stanford, Calif.: Stanford Graduate School of Business. https://www.gsb.stanford.edu/faculty-research/working-papers/interest-rate-risk-banking-0.

Drechsler, Itamar, Alexi Savov, Philipp Schnabl, and Olivier Wang. 2023. "Banking on Uninsured Deposits." Working Paper 31138. Cambridge, Mass.: National Bureau of Economic Research.

Duffie, Darrell. 2024. "Silicon Valley Bank and Beyond: Regulating for Liquidity." In *Getting Monetary Policy Back on Track*, edited by Michael D. Bordo, John H. Cochrane, and John B. Taylor. Stanford, Calif.: Hoover Institution Press.

Group of Thirty. 2024. *Bank Failures and Contagion: Lender of Last Resort, Liquidity, and Risk Management*. Washington: Author.

Hsu, Michael J. 2024. "Building Better Brakes for a Faster Financial World." Remarks at Columbia Law School, January 18. https://www.occ.gov/news-issuances/speeches/2024/pub-speech-2024-4.pdf.

Jiang, Erica Xuewei, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2023. "Monetary Tightening and U.S. Bank Fragility in 2023: Mark-to-Market Losses and Uninsured Depositor Runs?" Working Paper 31048. Cambridge, Mass.: National Bureau of Economic Research.

COMMENT BY

RAGHURAM RAJAN This is a thoughtful and important paper, consistent with the extraordinary caliber of the coauthors. It tees off the Silicon Valley Bank (SVB) crisis to argue that we need to reexamine bank regulations, especially given that banks themselves have evolved considerably over the last two decades. It suggests three important changes to regulations: (1) recognize that large banks are more like bond mutual funds in that they hold securities and make fewer information sensitive loans; authorities should consider stricter liquidity regulation for them but make access to liquidity easier by allowing pre-positioning of loans at the discount window; (2) require banks to recognize interest rate risk, at least in their "available for sale" securities portfolios if not their "held to maturity" securities portfolio; and (3) allow mergers of midsize banks because they have little franchise value.

I will comment on these proposals later, but first I want to ask three preliminary questions. One rationale for bank regulatory or supervisory intervention ex ante is to avoid externalities imposed by the banks (for instance, fire sales), which in turn might prompt liquidity or solvency bailouts by the authorities. So, first, how large were the externalities in the case of SVB and was a bailout of uninsured depositors once SVB experienced a run really required? Second, to what extent are bank activities a response to previous regulations, supervisory actions, and even monetary policy interventions? Third, is additional regulation necessary, and if so, where?

HOW LARGE ARE THE EXTERNALITIES FROM BANK FAILURE? The financing of loan-making intermediaries with short-term debt or with liabilities with extensive covenants is pervasive—shadow bank structures replicate bank structures in spirit if not in the details.[1] Whether the attraction is the cheap cost of issuing money-like liabilities that offer liquidity to holders or the discipline tough capital structures bring (not because depositors monitor

1. See, for example, Erel and Inozemtsev (2024).

but because they run at the first sign of trouble) or both, as in Diamond and Rajan (2001), does not really matter for the systemic negative externalities they may create down the line. But if discipline is the intent, repeated predictable bailouts privatize the gains from risk taking (the intermediary gains from the returns on the risks it takes) while socializing losses (the public bears the cost of paying off the liabilities the bank contracts at low cost). Before arguing for changing regulations, we must ask first if the problems in SVB were systemic, so much so that the Federal Deposit Insurance Corporation (FDIC) had to bail out its uninsured depositors.

Prima facie, it would seem that SVB management was either greedy or incompetent or both. The bank had around $57 billion in assets in 2018, and it grew to nearly four times that size by 2021, a period when the banking industry grew assets by only 29 percent (Barr 2023). A bank's spectacular growth is often an early warning sign of subsequent problems. The bank had significant investments in long-term securities even as its growth was financed by uninsured demand deposits. Regardless of whether it was searching for yield by investing the inflows in long-term securities, whether it believed its deposits would not reprice as the Federal Reserve raised interest rates (that is, its deposit betas were low), or whether it thought interest rates would stay low, SVB and its supervisors failed Risk Management 101. SVB was insolvent when its holdings were marked to market, a realization that triggered the run.

Clearly, the bailout was too late to stop the run. When SVB was taken over by the FDIC and its parts sold off to other banks, the direct losses as a consequence of the change in management may well have been small. There was no fire sale of individual assets. It is hard to imagine that tech firms did not obtain adequate service from their new bank. So the run seems to have imposed limited costs on the system because of the efficient transfer in ownership, something the FDIC has become adept at. If uninsured depositors had borne the full losses, they would have recovered 80–90 percent of their deposits according to Moody's, a painful lesson for depositing corporate treasurers on what it means to be uninsured but not necessarily debilitating for most.[2]

So why then did SVB's uninsured depositors have to be bailed out? Almost surely, the authorities feared contagion—that other banks were in a

2. Moody's Investors Service, "Moody's Downgrades SVB Financial Group (Senior Unsecured to C from Baa1) and Will Withdraw the Ratings," March 10, 2023, https://www.moodys.com/research/Moodys-downgrades-SVB-Financial-Group-senior-unsecured-to-C-from-Rating-Action--PR_474735.

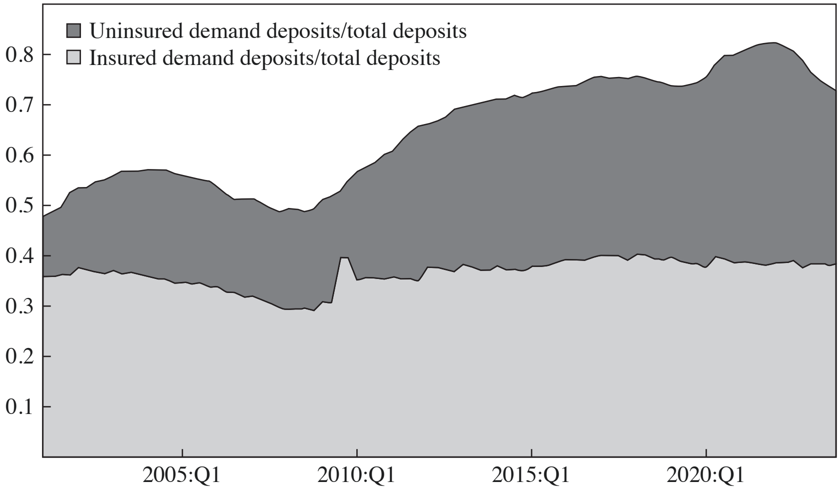
similar position of having long-term asset portfolios financed with uninsured demand deposits, and the losses sustained by the dramatic rise in interest rates made them subject to runs. Indeed, some twenty-two runs were under way (Cipriani, Eisenbach, and Kovner 2024). Jiang and others (2023) estimated that even if only half of uninsured depositors decided to withdraw in March 2023, almost 190 banks with assets of $300 billion were at a potential risk of insolvency. However, when they add going-concern franchise value to the mark-to-market value of assets, DeMarzo, Krishnamurthy, and Nagel (2024) find far fewer insolvent banks.

Nevertheless, the point is that "search for yield" behavior financed by uninsured demand deposits was widespread. Rational runs on the insolvent culprits may have resulted in painful losses for uninsured depositors but would not have been systemic. At the back of the banking authorities' minds, of course, is the worry that if they are not checked, runs may spread from the insolvent to the solvent—a full-fledged panic. Would letting SVB's uninsured depositors bear some losses have led to a full-fledged panic? We will never know because it was not allowed to happen, but it does raise the question of whether the authorities have the appetite for allowing depositor losses at any but the tiniest banks anymore. Undoubtedly, once the Treasury, the Fed, and the FDIC implicitly assured all uninsured depositors that the systemic risk exception invoked to bail out uninsured depositors in SVB and Signature Bank would be applied more widely, further bank runs stopped. At the same time, the authorities may have set a deeply problematic precedent for the future.

DID THE AUTHORITIES CONTRIBUTE TO BANK RISK? No matter how much regulators and supervisors emphasize principles-based regulation, in practice, rules matter because they give the supervisor safe harbor. Moreover, a principles-based supervisor may not have the political clout to highlight, and require remedial action on, vulnerabilities that are not traditional—until it has blown up, how do you know it will? Finally, after a crisis, the rules covering the most recently observed vulnerabilities are strengthened, and compliance is closely monitored—after all, at the very least, regulators and supervisors ought to close the stable door firmly, to show they are cognizant of the horse having bolted (Rajan 2009). Given all this, regulators and supervisors, like generals, tend to fight the last war vigorously.

In 2007–2008, the main issue was the credit risk buried in complex financial assets. There was little of all that in 2023, though there certainly were potential credit defaults in plain vanilla loans to commercial real estate. The biggest cause for concern was interest rate risk in long-term securities and loan portfolios, accentuated by deposit repricing and flight risk on the

Figure 1. The Composition of Domestic Demand Deposits in the United States



Source: Reproduced from Acharya and others (2024).

liability side, again possibly related to interest rates. When the Fed raised interest rates from June 2004, it did so steadily over a two-year period, with a predictable 25 basis point hike every meeting (Federal Reserve Board of Governors 2024). The rate hikes between March 2022 and July 2023 were much more rapid, with four 75 basis point hikes in succession. Moreover, during the period of quantitative easing (QE) preceding the rate hikes, as Acharya and others (2024) document (see figure 1), not only did the share of demand deposits to total domestic deposits go up from 60 percent in 2008 to 88 percent in 2021, the share of uninsured demand deposits to total domestic deposits went up from 24 percent to 47 percent. As a result of the change in deposit structure, which seems to have been little commented on by supervisors (Gopalan and Granja 2023) over the period of successive QEs, deposits became far less attached to the banks than in the past. Alert depositors rather than sleepy depositors dominated now, and deposits became more mobile. SVB was an outlier in this regard, but the phenomenon was more general.

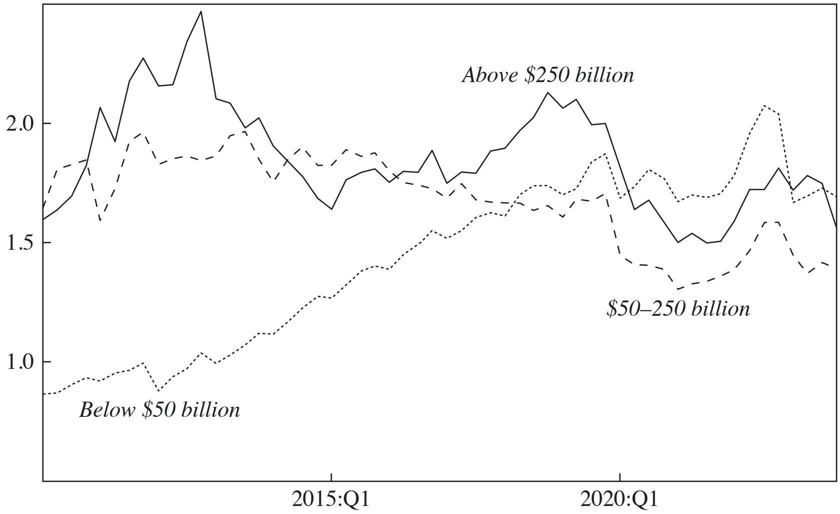
So rapid interest rate hikes were a double whammy for banks. They led to depressed long-term asset values of even safe assets and, simultaneously, to rapid repricing or flight of deposits in ways that banks were hitherto not used to. Supervisors did not anticipate rapid interest rate hikes (as the authors of this paper point out, "remarkably, the Federal Reserve's stress test scenarios in 2021 and 2022 did not include interest rate increases"), nor did they seem to recognize banks had become more vulnerable to rate hikes.[3] There is some evidence that supervisors noted some of SVB's failings, but they were not sufficiently concerned to press for a rapid response—so much so that SVB was allowed to unwind some of its interest rate hedges just before its demise (Levine 2023). The reason may well be that this was a very different scenario from the run-up to the global financial crisis (recall that with the onset of that crisis, the Fed cut rates, elevating the value of long-term securities). Perhaps supervisors' mindset had not shifted!

Another concern is the distribution of risks across the banking system. The authors show (in table 1 of the paper) that larger bank lending has fallen, and their securities holdings have increased substantially, with cash and short-term securities holdings going up by more than long-term securities holdings. Conversely, smaller banks have maintained their lending at a relative constant fraction of their assets, while their cash plus securities have fallen, especially in the cash plus short-term securities category. Prima facie, it would seem that smaller banks now have greater liquidity risk. Indeed, this is what Acharya and others (2024) show. Figure 2 from their paper suggests the sum of small banks' liquidity exposures (uninsured demand deposits plus lines of credit) to ready sources of liquidity (cash plus reserves plus repo-eligible securities) rose dramatically over the period of quantitative easing, from below 1 to peak at above 2 just before the Fed started raising interest rates. By contrast, for the largest banks, the ratio peaked before the liquidity coverage ratio (LCR) regulation was approved in 2014, and generally drifted down after. It was around 1.7 when the Fed started raising rates. While Acharya and others (2024) define small banks as those with below $50 billion in assets and large banks as those with above $250 billion in assets, table 1 is not inconsistent with their finding—the liquidity risk of smaller banks has gone up substantially in the last decade and a half. They have become more dependent for liquidity on other banks, the Fed, and the Federal Home Loan Banks. Indeed, it bears noting that SVB was one of these smaller banks before 2018.

It would then seem there is interesting specialization emerging within the banking system. The large banks are becoming contingent liquidity

3. DeMarzo, Krishnamurthy, and Nagel (2024) suggest that the excess holding of securities by banks whose franchise value had positive duration risk may have been supervisor-driven, with supervisors believing incorrectly that the securities were offsetting a negative duration franchise value. Gopalan and Granja (2023) show that bank supervisors started downgrading banks with substantial interest rate exposure only after the Fed started raising interest rates. Furthermore, they did not seem to recognize the risk posed by uninsured demand deposits, which Acharya and others (2024) show had spread through the system.

Figure 2. Claims to Potential Liquidity: (Credit Lines + Uninsured Demandable Deposits)/(Reserves + Eligible Assets) across Bank Size



Source: Reproduced from Acharya and others (2024).

Note: This figure plots the distribution across bank holding companies (BHCs) over time of claims to potential liquidity, which is the ratio of the sum of aggregate credit lines and demandable deposits to the sum of reserves and eligible assets between 2010:Q1 and 2023:Q4, with data (field) obtained for each component from Call Reports: Off-balance sheet unused loans or credit lines (RCFDJ457); Uninsured demandable deposits, obtained by subtracting time deposits of more than $250,000 ($100,000 before 2008:Q4) from total uninsured deposits, the latter being estimated from schedule RC-O of the Call Reports. Reserves reflect field RCFD0090, and eligible assets consist of Treasury and agency securities that were eligible for sale to the Fed for reserves in at least one quantitative easing round between 2008:Q4 and 2023:Q1. In particular, bank holdings of Treasury and agency securities are estimated as the sum of the bank's holdings of US Treasuries, obligations of US government agencies, and agency-backed mortgage-backed securities. The value of reserves and credit lines are set to zero if they are missing at the consolidated bank or bank holding company level for a given quarter. The ratio is aggregated by bank size categories. The size buckets are banks with assets above $250 billion, $50–250 billion, and below $50 billion in 2014:O3.

providers (Kashyap, Rajan, and Stein 2002), maintaining suitably liquid balance sheets, including substantial quantities of cash reserves and short-term securities, to do so. Perhaps the more complete application and enforcement of LCR regulation as well as higher capital requirements for systemically important banks forces them to move from holding loans on their balance sheet to using their balance sheet more contingently.[4] Small banks, in contrast, are in the more traditional business of relationship lending, with a significant portion of their assets still relationship loans. It is then particularly

4. Also see Erel and Inozemtsev (2024) on shifts in bank activities as a result of regulatory pressures.

worrying that these banks have increased their reliance on demandable deposits while shrinking their holdings of liquid assets (and even extending their maturity). From a systemic risk perspective, this change is most concerning as it creates a common exposure across many small banks.

The bottom line is that part of the reason small and midsize banks were collectively exposed to the risk of uninsured demandable claims in March 2023 was the prior Fed's balance sheet expansion and contraction. Why these banks did not hold more reserves and short-term securities, and why they instead lengthened the maturity of their securities portfolios (as in table 1) is not obvious. Perhaps it was classic bank search for yield, as Acharya and others (2024) imply. Regardless, the Fed's actions played a role in raising bank risk. Maybe the experience with these actions will lead the Fed to be more circumspect about using its balance sheet as a monetary policy instrument in the future. To the extent, however, that the Fed will continue to use its balance sheet in the future, it will have to consider the effects on the banking system and the potential need for regulations to offset adverse behavior.

MORE REGULATION? The possible adverse behavior engendered by future Fed policy has to be viewed with the additional knowledge that the SVB episode has enhanced the expectation that uninsured depositors in all banks above a (low) size threshold will be bailed out in the future in the event of a run. The authorities will have to worry that banks may have fewer qualms about financing with "cheap" uninsured demand deposits, rendered cheaper because of the anticipation the authorities will intervene. Furthermore, even if they do not run, uninsured depositors are less likely to be attached to the bank than traditional insured depositors and will be quicker to demand repricing. In other words, bank moral hazard and bank risk may increase as a consequence of SVB.

Where then to regulate? Apart from concern about whether supervisors can enforce a holistic mandate on risk taking, there is the important issue of risk migration. To the extent that certain entities are regulated or scrutinized more closely than others, risk migrates away from those entities but often ends up in less-scrutinized entities. So, for example, if liquidity positions are more closely scrutinized at large banks, liquidity risk moves to small banks or into the nonbank sector. Even though the authors emphasize that the costs of additional liquidity regulation may be lower at large banks—which seem more akin to money market bond funds—large banks were not the ones that got into trouble (large bank regulations did not apply to SVB because it was a small-to-midsize bank for much of the time when risks built up). Furthermore, the tighter liquidity regulation on large banks seems to have led them to draw reserves away from small banks. This would not be a problem if large banks were to lend liquidity freely in times of stress. Unfortunately, as Acharya and others (2024) argue, the residual source of liquidity for stressed banks seems to have been the Fed windows and the Federal Home Loan Banks, and not the large banks. In sum then, more uniform regulation—for instance, extending LCR to smaller banks—seems to them a more desirable first step than more regulations on the large banks. With these caveats, let us move to the specifics of the proposals.

PROPOSAL 1: PRE-POSITIONING LOANS AS COLLATERAL AT LARGE BANKS The SVB crisis suggests stricter liquidity regulation ought to be extended to smaller banks, as this paper commendably suggests. The most novel part of the proposal is to pre-position loans at the discount window, which is what I will focus on. This idea seems very sensible, addressing both the stigma associated with borrowing from Fed facilities (which tends to deter borrowing) as well as the possibility that the bank may have too few high-quality assets to raise secured funding quickly. The authors do a great job in raising concerns and addressing them. A few additional concerns are worth addressing.

First, if indeed large banks are moving to using their balance sheet contingently, offering liquidity to firms in case of need, they will be adding significant loans in times of aggregate liquidity stress. Of course, some of the large banks' liquidity needs could be offset by deposit inflows (Gatev and Strahan 2006). However, to ensure that large banks continue to intermediate liquidity, the central bank should lend against the new loans. These would not have been pre-positioned and may indeed be riskier than the norm—for instance, they will include drawdowns on lines of credit, which could be loans that no bank would make without having entered into a prior commitment. Shouldn't pre-positioning also include such contingent loans?

This leads to a second concern, which at the broadest level applies to all publicly provided insurance: the tendency to underprice it. The central issue in pricing is, of course, the haircut the Fed should apply on the value of these contingent loans to determine the amounts it lends. A related issue is the haircut it imposes on ordinary pre-positioned loans, knowing that the haircut is set in normal times while the Fed's liquidity is drawn upon in times of stress. Haircuts should anticipate such stress, but it will be hard for the central bank to get it right. Should the proposed haircuts be dynamic, increasing if conditions deteriorate more than anticipated? Dynamic haircuts would reduce the value of pre-positioning and may even set off a run if the haircut increases substantially, but it would allow the Fed to set lower haircuts up front given they can be changed.

Taking these considerations into account, perhaps the haircuts should be dynamic but change only after a lag—for instance, the Fed would reexamine the pre-positioned loans periodically, calculate the liquidity deficiency if the haircut reflected true risk of these loans, get the bank to fill the deficiency from other sources, and announce the new haircut after the next examination. The old haircut would prevail in between examinations, giving the bank the time between examinations to get its liquidity supply in order, while, of course, unavoidably exposing the Fed to more risk.

A final concern is that if the pre-positioning facility is available only to large banks, they may become even more attractive destinations for flight-to-safety money in times of stress. If they do lend the money back out, this is not entirely bad, given our earlier discussion. However, if they hoard it, then a liquidity facility available only to some banks may exacerbate the liquidity shortage in others. More generally, it is worth contemplating a liquidity facility that is widely available rather than one that is available only for some banks.

PROPOSAL 2: INTEREST RATE RISK A second proposal that is hard to argue with is a better treatment of interest rate risk exposures. Of course, a key concern is to get the overall exposure right—if only some part of the exposure is accounted for, the bank forced to recognize it, and the consequent valuation changes made to affect bank regulatory capital, the bank will try and manage down that exposure. If, however, that exposure is a hedge for other, harder-to-measure exposures, such as loan and deposit rate sensitivities, there is a risk that the bank could become overexposed to rate risk. DeMarzo, Krishnamurthy, and Nagel (2024) show how important it is to take all rate exposures into account and conclude that banks overall have positive duration. If so, an expanded securities portfolio can lead to greater interest rate risk, in which case requiring capital against the securities portfolio's interest exposure is a step in the right direction. However, such a conclusion will not be true for every bank.

Should we work toward a supervisory framework that tries to estimate the interest rate exposure of each bank's franchise and, following that, sets a fraction (from zero to 1) for the flow-through of securities portfolio valuation changes into capital? This would add to the complexity of the supervisors' task and expand discretion, with all the attendant previous caveats, but it may be better than mandating 100 percent pass-through. At any rate, I endorse the view that we need better understanding and treatment of bank interest rate exposures.

PROPOSAL 3: ALLOW MERGERS OF SMALL AND MIDSIZE BANKS The proposal to allow some mergers of banks below mega-bank size—so as to allow them to upscale from the midsize level that no longer seems to add value—once again makes sense. The concern is that a lot of small banks merge in this more liberal environment to become midsize banks—perhaps because small-bank managers have empire-building motives. After all, that is how we got some midsize banks in the first place. One possibility is to make it more attractive for banks to stay small and local. Can they get some of the benefits of scale without scaling up? Other countries have networks of small banks that offer mutual insurance, economies in purchasing technology, and some common resources—for example, the Rabobank network in the Netherlands. Should the impediments to such structures in the United States be identified and removed? Is it too far-fetched to imagine that some midsize banks might break up through management buyouts into such networks of small banks?

SUMMARY This is a great paper and has a number of interesting policy recommendations. Obviously, the analysis is intended to start a debate and will inspire more research. I am sure it will have that effect.

REFERENCES FOR THE RAJAN COMMENT

Acharya, Viral V., Rahul S. Chauhan, Raghuram G. Rajan, and Sascha Steffen. 2024. “Liquidity Dependence and the Waxing and Waning of Central Bank Balance Sheets.” Working Paper 31050. Cambridge, Mass.: National Bureau of Economic Research.

Barr, Michael S. 2023. *Review of the Federal Reserve’s Supervision and Regulation of Silicon Valley Bank*. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/publications/files/svb-review-20230428.pdf.

Cipriani, Marco, Thomas M. Eisenbach, and Anna Kovner. 2024. “Tracing Bank Runs in Real Time.” Staff Reports 1104. New York: Federal Reserve Bank of New York. https://www.newyorkfed.org/research/staff_reports/sr1104.

DeMarzo, Peter, Arvind Krishnamurthy, and Stefan Nagel. 2024. “Interest Rate Risk in Banking.” Working Paper. https://bpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/f/575/files/2024/06/Interest_Rate_Risk_in_Banking.pdf.

Diamond, Douglas W., and Raghuram G. Rajan. 2001. “Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking.” *Journal of Political Economy* 109, no. 2: 287–327.

Erel, Isil, and Eduard Inozemtsev. 2024. “Evolution of Debt Financing toward Less-Regulated Financial Intermediaries in the United States.” Working Paper 32114. Cambridge, Mass.: National Bureau of Economic Research.

Federal Reserve Board of Governors. 2024. “Open Market Operations.” July 10. https://www.federalreserve.gov/monetarypolicy/openmarket.htm.

Gatev, Evan, and Philip E. Strahan. 2006. “Banks’ Advantage in Hedging Liquidity Risk: Theory and Evidence from the Commercial Paper Market.” *Journal of Finance* 61, no. 2: 867–92.

Gopalan, Yadav, and João Granja. 2023. “How (In)effective Was Bank Supervision during the 2022 Monetary Tightening?” Working Paper 2023-130. Chicago: Becker Friedman Institute for Economics at the University of Chicago.

Jiang, Erica Xuewei, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2023. "Monetary Tightening and U.S. Bank Fragility in 2023: Mark-to-Market Losses and Uninsured Depositor Runs?" Working Paper 31048. Cambridge, Mass.: National Bureau of Economic Research.

Kashyap, Anil K., Raghuram G. Rajan, and Jeremy C. Stein. 2002. "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking." *Journal of Finance* 57, no. 1: 33–73.

Levine, Matt. 2023. "It's Never Too Late for Banks to Hedge." Bloomberg, May 9. https://www.bloomberg.com/opinion/articles/2023-05-09/it-s-never-too-late-for-banks-to-hedge?sref=cY1UiIYj.

Rajan, Raghuram G. 2009. "The Credit Crisis and Cycle-Proof Regulation." *Federal Reserve Bank of St. Louis Review* 91, no. 5, pt. 1: 397–402.

GENERAL DISCUSSION Donald Kohn questioned what drove the growth of uninsured deposits. He observed that while uninsured deposits as a percentage of domestic deposits seemed to follow a trend in the 1990s and early 2000s, in the following decades, they occasionally stair-stepped and abruptly increased. He theorized that this might be a result of the zero interest rates at those times, which complicated banks' time deposit and demand deposit mix. Banks may have been reluctant to price demand deposits at a negative interest rate because they did not want to lose business. This could make demand deposits more appealing to depositors, meaning that the relative increase in demand deposits might come from the depositor side rather than from the banks seeking demand deposits.

Wendy Edelberg presented an alternative theory of what drove the rising share of deposits held as demand deposits. She asserted that since the US Treasury cannot change the amount of money in the banking system, and since federal borrowing from abroad is modest compared to the overall increase in federal borrowing, the increase in demand deposits might be primarily explained by quantitative easing (QE). Edelberg then suggested that monetary policymakers in the future should account for how QE would affect demand deposits. Even so, she acknowledged that regulators could not practically vary the deposit insurance level every quarter, so they could not simply raise the required deposit insurance level whenever there was higher QE.

Randall Kroszner pointed out that since the United States is currently experiencing quantitative tightening (QT), by the end of 2024, there should be evidence for whether or not QE is driving deposits. If deposits go down very significantly under QT, the hypothesis that QE drives deposits would have more support.

Laura Nicolae contended that QE's role in driving long-term deposit growth was not totally clear. She commented that the Federal Reserve buying bonds from a bank does not necessarily create deposits—it just switches out reserves for bonds on a bank's balance sheet. Similarly, the Federal Reserve buying bonds from a nonbank might create deposits in the short term but does not guarantee that those deposits will remain where they are. Nicolae instead attributed the long-term growth in deposits to growth in demand for deposits, noting that figure 1 (panel B) in the paper shows that deposits scale with wealth.[1] She also noted that deposits have been growing for decades, through periods of both QE and QT.

Andrew Atkeson reflected on his time at the Federal Reserve Board's Model Validation Council, an advisory body that provides guidance for the models used in bank stress tests. He claimed that the issues of interest rate risk and run risk that led to the collapse of Silicon Valley Bank (SVB) were never considered, despite the purpose of a stress test being to anticipate new and evolving risks. He wondered if there were institutional changes that could be made to allow regulators to address future, otherwise unanticipated, problems.

Laurence Ball echoed the paper's sentiment that stress tests and regulators have made overly optimistic assumptions about the risk presented by uninsured deposits. Referring to one of his own papers, he observed that the current stress test scenario makes additional problematic assumptions.[2] According to Ball, outflow rates for repurchase agreement financing are too low in the current scenario. Furthermore, Ball posited that the scenario's worst error was assuming that banks' outflows are largely offset by inflows, which would come from cutting off financing to customers. Banks might be loath to do this because it would destroy their ability to do business in the future.

Although Ball advocated for stronger liquidity coverage ratio (LCR) regulations, he granted that regulators might conclude that banks ought to hold so much liquidity that there would be none left over for lending. In light of that potentiality, he supported the idea of pre-positioning collateral at the Federal Reserve's discount window. He argued the LCR rule was meant to allow banks to survive crises without borrowing, which might be an unrealistic demand. Future scenarios should make the Federal Reserve's role more explicit and incorporate the option of borrowing.

Burcu Duygan-Bump contemplated the role of liquidity regulations compared to that of the Federal Reserve as the lender of last resort (LOLR).

1. Here refers to figure 1 (panel B) in the conference draft of the paper, available at: https://www.brookings.edu/events/bpea-spring-2024-conference/.

2. Laurence Ball, "Liquidity Risk at Large U.S. Banks," *Journal of Law, Finance, and Accounting* 7, no. 2 (2023): 229–72.

She questioned if a bank's pre-positioned collateral needed to be counted toward its LCR and if that presented any dissonance. She thought not but believed the role of LCR versus LOLR merited further consideration.

Turning now toward the discussion of how to manage interest rate risk and liquidity risk, Kroszner questioned why the United States did not require a capital charge whenever banks take on interest rate risk. This rule had been adopted by many other countries, and Kroszner emphasized that no other country experienced similar issues during the collapse of SVB. He clarified that Credit Suisse's collapse, while contemporaneous, had nothing to do with interest rate risk.

Kohn concurred with the paper's points about penalizing banks that use hold-to-maturity accounting to avoid acknowledging unrealized changes in the market value of their held bonds—called mark-to-market gains and losses. He affirmed that extending the mark-to-market was a vital part of managing interest rate risk and liquidity risk.

Andrew Fieldhouse remarked that conversations with a colleague from regional reserve bank had led him to believe that SVB's collapse was primarily due to duration risk on agency mortgage-backed securities (MBS) rather than interest rate risks on Treasuries. He also recommended that the authors consider the feasibility of hedging refinancing risk on agency MBS in addition to their examination of interest rate risk.

Samuel Hanson compared banks to insurers, saying that insurers marking their assets while not marking their liabilities to market would be economically incoherent. While challenging, Hanson contended that requiring banks to mark both sides of their balance sheet was clearly necessary. He went on to agree with Arvind Krishnamurthy's discussion about creating a capital charge that scaled with the amount of liquidity transformation, since more liquidity transformation meant greater risk of loss of franchise value.

The discussion also touched on why banks seem so hesitant to use the Federal Reserve's discount window. Kroszner expressed concern that the discount window was poorly run and not that user-friendly, and argued that these issues, as well as the stigma associated with using the discount window, must be addressed if the Federal Reserve wanted more banks to use the window more regularly.

Kohn pointed out that banks went to the Federal Home Loan Banks (FHLBs) rather than the discount window because it was cheaper overall. He mentioned that there had been recent discussion of reforming the FHLBs to refocus them on only supporting the mortgage market. Kohn's recommendation to address the stigma around the discount window was for regulators to acknowledge that discount window access can be a part of a recovery and resolution plan by allowing discount window utilization in stress tests. Additionally, the penalty for using the discount window should be reduced. According to Kohn, the United Kingdom and the eurozone had successfully reduced the stigma by moving to reserve management processes under which banks borrow whenever reserves get sufficiently scarce.

Responding to Jeremy Stein's presentation, Robert Hall drew a sharp distinction between banks and bond mutual funds, noting that it was incorrect to claim banks were becoming mutual funds. Hall said that mutual funds are organized under the Investment Company Act of 1940. This makes them run proof, since someone withdrawing from a fund is paid off immediately at market value. In contrast, banks are immensely at risk of runs, making turning banks into a mutual fund highly desirable.

Stein acknowledged the critique by Hall. He observed that while banks are increasingly resembling MBS bond funds on the asset side, they are much less suited to handle duration risk on the liability side when issuing uninsured deposits. All else being equal, this makes it better for MBS to be held in a bond fund than by a bank that finances the MBS with uninsured deposits. An ideal regulatory framework would level the playing field without encouraging or disadvantaging the bond funds relative to the banks. Unfortunately, bailing out uninsured depositors creates an implicit subsidy, which encourages MBS to be held in banks.

Jón Steinsson discussed the two socially valuable contributions of banks that the paper highlighted: the provision of information-intensive loans and the provision of transaction services. He remarked that the latter service is potentially overlooked. Payment systems work so well today that people forget how crucial they are to a well-functioning economy. He believed this should not be taken for granted, and he pointed out that the debate surrounding which of these two socially valuable contributions to emphasize was often extreme and binary with one side favoring narrow banking and the other side hostile to any increase in capital requirements. He felt that the results of the paper should tilt this debate in the direction of higher capital requirements being optimal.

Steinsson also emphasized another service provided by banks—the creation of liquidity. Steinsson described how banks are traditionally able to make long-term assets very liquid. He worried that the paper's focus on having banks hold short-term assets against uninsured deposits would mean that there would be no institutions left to hold long-term assets.

Adi Sunderam commented that Steinsson—and almost everyone else, including the other authors—seemed to take it as a given that it is always socially valuable to accommodate liquidity demand and safe asset demand.

Sunderam contested this assumption, as accommodating these demands could be negative under some circumstances, such as providing liquidity that facilitated crime.

Kroszner cautioned the authors about being too cavalier about creating regulatory systems that push banks out of information-intensive lending. He argued that just because banks are not doing much information-intensive lending does not mean it is a good idea to push them out entirely—what little they are currently doing might still be very socially valuable. He referred to a recent Bank for International Settlements paper, which found that, following a crisis, lending and investment fell and stayed down at firms that were primarily reliant on nonbanks.[3]

Stein reassured Kroszner, indicating that the authors were being careful not to push banks out of providing loans entirely, as shown by their proposal to allow loans to be pre-positioned at the discount window. Although the authors specifically hoped to lean against banks holding too many long-term MBS, they were careful not to recommend changes that were too extreme because banks only holding short-term securities could create equilibrium issues.

Şebnem Kalemli-Özcan considered whether banks of different sizes should be subject to the same regulations. Kalemli-Özcan inquired why the authors did not apply the LCR to all banks, including banks with assets below $50 billion. She suggested that extending LCR regulations to all banks might trigger desirable endogenous mergers, cutting down on a significant number of banks in the United States. She observed from the Federal Reserve's stress test results (FR Y-14) that small and midsize enterprises in fact borrow mainly from medium-size and large banks.[4] This means that reducing the number of small banks might not cause issues for small businesses.

Anna Paulson brought up the fact that larger and smaller banks have different business models. Consequently, one-size-fits-all regulations create different externalities. For instance, large banks that have fire sales often only part with less information-intensive assets compared to smaller banks. She identified this as a reason why creating blanket regulation was difficult.

Hanson agreed with the discussants and conference participants that applying liquidity regulations to smaller banks was worth considering.

3. Iñaki Aldasoro, Sebastian Doerr, and Haonan Zhou, "Non-bank Lending during Crises," BIS working papers 1074 (Basel: Bank for International Settlements, 2023).

4. Cecilia R. Caglio, R. Matthew Darst, and Şebnem Kalemli-Özcan, "Collateral Heterogeneity and Monetary Policy Transmission: Evidence from Loans to SMEs and Large Firms," working paper 28685 (Cambridge, Mass.: National Bureau of Economic Research, 2024).

He then answered Kalemli-Özcan's question, explaining that the paper anchored at banks with $100 billion in assets because that was where a 2018 law that amended the Dodd-Frank was set. This meant that $100 billion was the lowest level that the Federal Reserve could regulate without requiring further action from Congress. Sunderam added that the paper focuses on midsize banks because of what happened with SVB, but he theorized that the logic of the model should extend.